PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-030228

(43)Date of publication of application : 31.01.2003

(51)Int.Cl. G06F 17/30

(21)Application number : 2001-212555

(22)Date of filing : 12.07.2001

(71)Applicant : CASIO COMPUT CO LTD

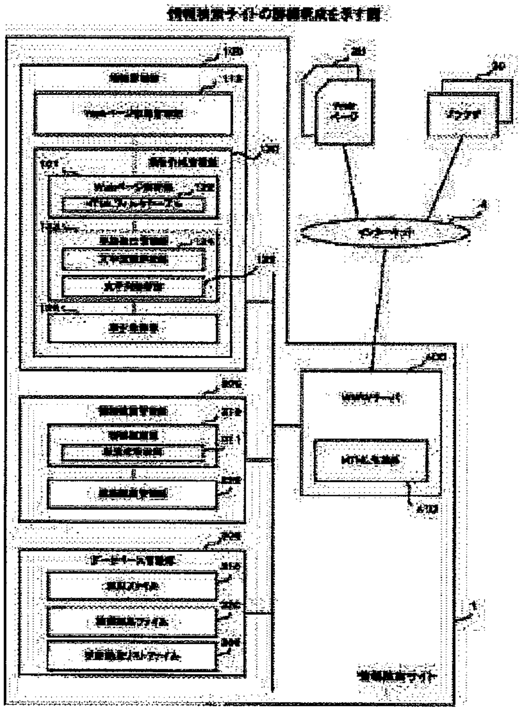
(72)Inventor : TERADA TOSHIHITO

(54) SYSTEM AND METHOD FOR RETRIEVING INFORMATION, AND PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To make information retrieval results provided by a retrieval engine to be more appropriate to a retrieval target of an information retriever.

SOLUTION: A character enhancement analyzing part 124 acquires an attribute showing enhancement given to a character string included in a web page 20. An index registering part 126 associates logical position information of the web page 20 and an enhancement attribute about a character string being an extraction source of a word with the word extracted from the character string by a character string analyzing part 125 to register them in an index file 310. An information retrieving part 210 acquires position information associated with a word representing a retrieval target. A retrieval result managing part 220 sorts the acquired position information to make a word with an enhancement attribute associated to take precedence with respect to a word made to be an object of retrieval of the position information. An HTML preparing part 410 prepares an HTML file representing the sorted position information and transmits the HTML file to the Internet 4.



* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Claim(s)]

[Claim 1]A word contained in document information characterized by comprising the following currently released on a communication network, An index file which matches position information which shows a document information position in which it is the information which shows a logical position on this communication network, and information containing this word exists is prepared, A system which presents position information corresponding to a word which searches this index file based on a word showing an object of search, and expresses this retrieval object.

An emphasis-attributes acquisition means which acquires emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis.

An extraction means to extract a word from said character string.

A registration means to match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word with a word extracted by said extraction means, and to register with said index file.

A search means to acquire position information which searches said index file based on a word showing said retrieval object, and is matched with this word from this index file, A presenting means which gives priority to that by which said emphasis attributes are matched with a word which this search means made an object of search of this position information in said index file among position information acquired by said search means, and presents this position information.

[Claim 2]The information retrieval system according to claim 1, wherein an attribute given to a character string contained in said document information is an attribute which shows color of a character used in order to display an attribute which shows a size of a character used when displaying this character string, or this character string.

[Claim 3]The information retrieval system comprising according to claim 1 or 2:

A frequency-of-occurrence calculating means which computes the frequency of occurrence in this document information about an attribute given to a character string by which said emphasis-attributes acquisition means is included in said document information for this every attribute.

An emphasis-attributes setting-out means to set up a standard which distinguishes whether said attributes are said emphasis attributes based on the frequency of occurrence for this every attribute, and an emphasis-attributes discriminating means which distinguishes whether attributes given to a character string contained in said document information based on said standard are said emphasis attributes.

[Claim 4]the information retrieval system according to claim 1 which is rich and is characterized by a thing which is said emphasis attributes about this attribute, and to make when said emphasis-attributes acquisition means shows that a character used in order that an attribute given to a character string contained in said document information may display this character string is made into a bold letter.

[Claim 5]Inside of position information from which said presenting means was acquired by said search means, The information retrieval system according to claim 1 giving priority to a thing which has many number of said emphasis attributes matched with a word which this search means made an object of search of this position information in said search file, and showing this position information.

[Claim 6]An index file which matches position information which shows a document information position characterized by comprising the following in which information exists is prepared, It is the method of showing position information corresponding to a word which searches this index file based on a word showing an object of search, and expresses this retrieval object, Emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis are acquired, Match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word with a word which extracted a word from said character string and was extracted from said character string, and it registers with said index file, Position information which searches said index file based on a word showing said retrieval object, and is matched with this word is acquired from this index file, An information retrieval method characterized by what priority is given to that by which said emphasis attributes are matched with a word made into an object of search of this position information in said search file among position information acquired by said search, and this position information is shown for.

A word contained in document information currently released on a communication network.

It is the information which shows a logical position on this communication network, and is this word.

[Claim 7]By performing a computer, Processing which prepares an index file which matches position information which shows a document information position in which it is the information which shows a logical position on a word contained in document information currently released on a communication network, and this communication network, and information containing this word exists, It is a program for making processing which presents position information corresponding to a word which searches this index file based on a word showing an object of search, and expresses this retrieval object perform to this computer, Processing which acquires emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis, Processing which matches processing which extracts a word from said character string, and said emphasis attributes given to a character string of said position information [about this word], and extraction origin of this word at a word extracted from said character string, and is registered into said index file, Inside of processing which acquires position information which searches said index file based on a word showing said retrieval object, and is matched with this word from this index file, and position information acquired by said search, A program for making processing which gives priority to that by which said emphasis attributes are matched with a word made into an object of search of this position information in said search file, and presents this position information perform to a computer.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Detailed Description of the Invention]
[0001]
[Field of the Invention]This invention relates to the art of enabling it to provide the information which agreed appropriately by the demand, to the demand of search especially about the art of retrieving information.
[0002]
[Description of the Prior Art]In recent years, the number of the Web pages provided by the WWW (WorldWideWeb) system on the Internet is continuing increasing explosively by the spread of the Internet. On the Internet, many search engines which provide the service which retrieves the information made into the purpose out of this huge information are established.
[0003]There are some which are called the robot type as one of the methods which collects the information on a network of a search engine. In a robot type search engine, the robot program called a spider or a crawler is started periodically, Automatic collection of the HTML (HyperText Markup Language) file expressing the Web page currently exhibited on the Internet is performed. When information retrieval is performed and the information retrieval person using a search engine gives a closely related keyword to the target information at a search engine site, Processing which extracts that in which the keyword was contained from the collected file is performed, and an information retrieval person is provided with the list of Web pages which are the keyword and which are contained as search results with the information which shows the logical position on the Internet about the Web page.
[0004]
[Problem(s) to be Solved by the Invention]Generally, since the robot type search engine is performing [no] automatically processings of a to [from collection of information / offer of search results] by computer and operation of the information by judgment of human being intervenes there, The arrangement about the quality of the genre to which the collected information belongs, or its information is not made. Therefore, if search by coincidence of a mere keyword was performed on the occasion of search of information, a Web page including

important information is buried in search results, or. Or there were not few cases where it will be mostly contained in search results only in about the Web page in which what is called a search noise, i.e., the low information on usefulness, is contained.

[0005]Making more suitable the result of the information retrieval which a search engine provides in view of the above problem to an information retrieval person's retrieval object is the issue which this invention tends to solve.

[0006]

[Means for Solving the Problem]A word contained in document information to which this invention is opened on a communication network, An index file which matches position information which shows a document information position in which it is the information which shows a logical position on this communication network, and information containing this word exists is prepared, It is premised on a system or a method of showing position information corresponding to a word which searches this index file based on a word showing an object of search, and expresses this retrieval object.

[0007]And an information retrieval system which is one of the modes of this invention, An emphasis-attributes acquisition means which acquires emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis, In a word extracted by extraction means to extract a word from said character string, and said extraction means. A registration means to match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word, and to register with said index file, A search means to acquire position information which searches said index file based on a word showing said retrieval object, and is matched with this word from this index file, A presenting means which gives priority to that by which said emphasis attributes are matched with a word which this search means made an object of search of this position information in said index file among position information acquired by said search means, and presents this position information, SUBJECT mentioned above is solved by constituting so that it may ****.

[0008]An attribute given to a character string contained in said document information here is an attribute which shows color of a character used in order to display an attribute which shows a size of a character used when displaying this character string, for example, or this character string. It is possible that document information containing a word emphasis is instructed to be has a high possibility that information that importance is high is included about the word. Therefore, when two or more document information corresponding to a word which shows a search condition is opened to a communication network according to composition mentioned above, Since the word is emphasized, and priority is given to position information about document information considered that importance is high, it makes and it is shown, a result of information retrieval will become more suitable to an information retrieval person's retrieval object.

[0009]In an information retrieval system concerning this invention mentioned above, said emphasis-attributes acquisition means, A frequency-of-occurrence calculating means which computes the frequency of occurrence in this document information about an attribute given to a character string contained in said document information for this every attribute, An emphasis-attributes setting-out means to set up a standard which distinguishes whether said attributes are said emphasis attributes based on the frequency of occurrence for this every attribute, It may constitute so that it may have an emphasis-attributes discriminating means which distinguishes whether attributes given to a character string contained in said document information are said emphasis attributes based on said standard.

[0010]Since it can judge now that a character string to which a unique attribute is given in document information is emphasized in that document information according to this composition, this document information can be registered into an index file for a word contained in this character string as what has high importance.

[0011]In an information retrieval system concerning this invention mentioned above, said emphasis-attributes acquisition means, When making into a bold letter a character used in order that an attribute given to a character string contained in said document information may display this character string is shown, it may be made to consider that these attributes are said emphasis attributes.

[0012]According to this composition, in order to show emphasis of a character string in document information, the attribute of a purport that a display by a bold letter currently performed widely is performed can be promptly judged with emphasis attributes. In an information retrieval system concerning this invention mentioned above, said presenting means, Priority is given to a thing which has many number of said emphasis attributes matched with a word which this search means made an object of search of this position information in said search file among position information acquired by said search means, and it may be made to show this position information.

[0013]According to this composition, for a word by which a thing which has many number of emphasis attributes given to a certain character string is contained in this character string, this document information can be registered now into an index file as what has higher importance, and a result of information retrieval will become still more suitable to an information retrieval person's retrieval object.

[0014]An information retrieval method which is one of the another modes of this invention, Emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis are acquired, Match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word with a word which extracted a word from said character string and was extracted from said character string, and it registers with said index file, Position information which searches said index file based on a word showing said retrieval object, and is matched with this word is acquired from this index file, The same operation and effect as an information retrieval system concerning this invention mentioned above are acquired by giving priority to that by which said emphasis attributes are matched with a word made into an object of search of this position information in said search file among position information acquired by said search, and showing this position information.

[0015]SUBJECT mentioned above by making a computer execute the program also in a program for making processing which consists of the same procedure as an information retrieval method concerning this invention mentioned above perform to a computer is solvable.

[0016]

[Embodiment of the Invention]Hereafter, an embodiment of the invention is described based on a drawing. Drawing 1 is a figure in which the information retrieval site which carries out this invention shows the entire configuration of the communication network which provides an information search service.

[0017]In drawing 1, the Internet 4 which is all a communication network is accessed, and data can be delivered [the information retrieval site 1 the offer-of-information site 2a, 2b, 2c, 2d, and the user terminals 3a and 3b] and received mutually. The information retrieval site 1 is a WWW server system which provides a robot search type information search service for the user terminals 3a and 3b, is provided with the Research and Data Processing Department 100, the information retrieval Management Department 200, the information database Management Department 300, and the WWW server Management Department 400, and is constituted.

[0018]The Research and Data Processing Department 100 performs automatic collection of the information currently released on the Internet 4, and accumulates the collected information in the information database Management Department 300. The information retrieval Management Department 200 retrieves the information accumulated in the information database Management Department 300 according to the demand of the information retrieval sent via the Internet 4, and returns the result of the search to a requiring agency.

[0019]At the information database Management Department 300, accumulation of the information collected by the Research and Data Processing Department 100 and search of the information by the information retrieval Management Department 200 are performed. The processing by which the WWW server part 400 transmits the collected information which is sent via the Internet 4 to the Research and Data Processing Department 100, Processing which transmits the demand of the information retrieval sent via the Internet 4 to the information retrieval Management Department 200, and processing of sending out of a Web page in which the information which shows the result of the information retrieval sent from the information retrieval Management Department 200 is expressed are performed.

[0020]The offer-of-information site 2a, 2b, and 2c and 2d are WWW server systems which exhibit Web pages 20a, 20b, 20c, and 20d on the Internet 4, respectively. Although four offer-of-information sites are shown in drawing 1, the number of the offer-of-information sites connected to the Internet 4 may be arbitrary.

[0021]The user terminals 3a and 3b, respectively The offer-of-information site 2a, 2b, 2c, And it is a computer which can perform the browsers 30a and 30b which are the software which peruses the Web page provided from 2 d or the information retrieval site 1, and is operated by the information retrieval person who requests search of the information currently released on the Internet 4 to the information retrieval site 1. Although two users are shown in drawing 1, the number of the user terminals connected to the Internet 4 may also be arbitrary.

[0022]These information retrieval sites 1, the offer-of-information site 2a, 2b, 2c and 2d, and the user terminals 3a and 3b, The computer by which all have standard composition, i.e., CPU which controls each component by executing a control program, The storage parts store used as the work area at the time of the memory and CPU of a control program which consist of a ROM, RAM, a magnetic storage device, etc., and make CPU control each component executing a control program, or a storage area of various data, It can also constitute using a computer provided with the input part from which various kinds of data corresponding to operation by a user is acquired, the outputting part which show a display etc. various kinds of data and of which a user is notified, and the I/F part which provides the interface function for connecting with a network.

[0023]Next, drawing 2 is explained. The figure shows the detailed composition of the information retrieval site 1 in drawing 1 which carries out this invention. As shown in drawing 2, the Research and Data Processing Department 100 has the Web page collection Management Department 110 and the index production part 120, and is constituted, The information management retrieval part 200 is provided with the information retrieval section 210 and the search-results Management Department 220, and is constituted, The database manager 300 has the index file 310, the search-results file 320, and the search-results list file 330, and is constituted, and WWW server 400 is provided with the HTML preparing part 410, and is constituted.

[0024]The Web page collection Management Department 110 performs periodically automatic collection of Web page 20 currently exhibited on the Internet 4. The position information on Web page 20 by which the index build Management Department 120 was collected by the Web page collection Management Department 110, That is, the record used as the index which can lengthen the position information which shows the logical position on the Internet 4 in which Web page 20 exists is created, and it registers with the index file 310. The index build Management Department 120 has the Web page analyzing parts 121, the word extraction Management Department 123, and the index registering part 126, and is constituted.

[0025]The Web page analyzing parts 121 create the HTML filter table 122 which makes the unit of a record each HTML tag described by the text of the HTML file which analyzes Web page 20 and is expressing Web page 20. processing which analyzes the conditions of the character style which is emphasized when they are displayed as a screen of Web page 20 in the character string shown in the HTML filter table 122, and which can be been [character style / it] rich and made being performed by the character emphasis analyzing parts 124, and at the word extraction Management Department 123, Analysis of the character string shown in the HTML filter table 122 is conducted by the character string analyzing parts 125, and a word is extracted from the character string.

[0026]The logical position information [the index registering part 126 makes an entry the word extracted by the character string analyzing parts 125 and / entry / the] on the Internet 4 about Web page 20, The index record in which the abstract of Web page 20 in which the word was contained, and the form set as the word by Web page 20 matched the attribute flag which shows that it agrees on the conditions of the character style obtained in the analysis in the character emphasis analyzing parts 124 is registered into the index file 310.

[0027]The information retrieval section 210 acquires the demand of the information retrieval sent from the user terminal by control of the browser 30 currently performed with one of the user terminals connected to the Internet 4 from the WWW server part 400, The search formula showing the conditions of the information retrieval is taken out from the demand, and it stores in the search formula storage 211. And the word (keyword) which searches the index file 310 and is shown in the search formula acquires the index record used as a title, and stores in the search-results file 320.

[0028]If search by the information retrieval section 210 is completed, the search-results Management Department 220, The position information and the abstract which are shown in the index data stored in the search-results file 320, and the total number of the attribute flag attached with the index record corresponding to the position information are stored in the search-results list file 330. And the position information stored in the search-results list file 330 is sorted according to the total number.

[0029]The HTML preparing part 410 creates the HTML file expressing the Web page which receives the search-results list which consists of sorted position information which is stored in the search-results list file 330 and as which the search-results list is expressed. The created HTML file is addressed to the user terminal in which the browser 30 is performed, and is sent out to the Internet 4 by the WWW server part 400.

[0030]Next, the details of processing of the collection of a Web page performed in the Research and Data Processing Department 100 which the information retrieval site 1 has, and generation of an index are explained. Drawing 3 shows the example of Web page 20 which is opened to the Internet 4 and collected by the information retrieval site 1. In the figure, the browser's 30 inspection of the HTML sauce shown in (b) will display the screen shown in the figure (a).

[0031]Drawing 4 is explained here. The figure is a flow chart which shows the contents of processing of the index production processing performed in the Research and Data Processing Department 100. By performing this processing, collection of a Web page and generation of an index are performed in the Research and Data Processing Department 100. First, in S101, only when it was distinguished at the Web page collection Management Department 110 whether the present date is the collection designated date of Web page 20 specified beforehand, this decision result is set to Yes and the present becomes that designated date, processing progresses to S102. Although the method of specification of this date is arbitrary, specification called the monthly final day of month end, etc. is performed, for example.

[0032]In S102, processing of a round and collection of Web page 20 currently exhibited on the Internet 4 is performed by the Web page collection Management Department 110. The technique of this round and collection should just use as it is what is performed from the former by the well-known robot type search engine.

[0033]In the Web page analyzing parts 121, the initial value 1 is substituted for the variable m used as page pointers which are pointers for pointing at a time to 1 page of Web pages 20 of a large number collected at the front step S103. The structure of the page to which it points by the current value of the page pointers m in Web page 20 collected by processing of S102 in S104 is analyzed by the Web page analyzing parts 121, The HTML filter table 122 is generated by the Web page analyzing parts 121 in S105 continuing.

[0034]The HTML filter table generated from the Web page shown in drawing 3 is shown in drawing 5. Analysis of the HTML sauce shown in drawing 3 (b) by the Web page analyzing parts 121 will generate the HTML filter table shown in drawing 5. When it explains further, referring to drawing 3 (b) for the contents of processing of S104, in the Web page analyzing parts 121. The text of the HTML sauce of an analytical object, i.e., the portion pinched between the start tag of <BODY> and the end tag, is made into the object of analysis, and the structure of each sentence where the <BR> tag (line feed tag) in the portion is considered as a pause of a sentence, and is contained in the text is analyzed.

[0035]If signs that the HTML filter table shown in drawing 5 from the HTML sauce shown in drawing 3 (b) is created are explained, First, let the portion pinched between the start tag of the <BODY> tag and end tag which are the description parts of the text in HTML sauce, i.e., the portion put between the <BODY> tag and the </BODY> tag, be an object of the analysis by processing of S104.

[0036]here -- first -- an analytical object -- a portion -- it can set -- the beginning -- < -- BR -- > -- a tag -- describing -- having -- **** -- a part -- up to -- a portion -- namely, -- "-- < -- FONT SIZE -- = -- " -- six -- " -- COLOR -- = -- " -- # -- FF -- 0000 -- " -- > -- < -- B -- > -- easy -- cooking -- < -- /-- B -- > -- < -- /-- FONT -- > -- " -- a portion is analyzed. <FONT SIZE="6"COLOR="#FF0000"> Here the becoming tag, the "easy dish" described by the portion pinched between the start tag of <FONT>, and the end tag -- a character string -- character size -- "6" -- considering it as a size -- and a character color -- "#FF0000" -- what is displayed as a color shown numerically is shown. The color shown for this figure is red.

[0037]Displaying the character string "the easy dish" <B> [which is described by the portion by which the becoming tag is sandwiched between the start tag of <FONT> and the end tag] Becoming in a bold letter is shown. The analysis processing by S104 conducts analysis mentioned above, and the record shown in the 1st line of drawing 5 which means the contents of this analysis result is stored in the HTML page filter 122 by processing of S105 performed next. An "easy dish" is stored in the column of a "character string" when this record is explained, "STRING" which shows that an "easy dish" is a character string is stored in the column of "classification", and the attribute which shows that "6" and a "color" are displayed for "character size" as "red" about the character string of an "easy dish" is stored in the column of a "character string attribute." By furthermore storing the flag "1" in the column of the "bold letter" in a "character string attribute" shows that displaying the character string of an "easy dish" in a bold letter was shown. And the record in which existence of the <BR> tag which only makes it the contents for "classification" to be "BR" after the is shown is stored in the 2nd line of the HTML page filter 122.

[0038]Analysis of the remaining portion in the text part of the HTML sauce shown in drawing 3 (b) is conducted similarly hereafter, and the HTML page filter 122 shown in drawing 5 in this way is generated. Although the SIZE attribute in the <FONT> tag may be shown like "+1", "-1", etc. by the relative value over the usual displayed character size, When such, the relative value is registered as character size, and about the character string by which specification of character size is not made, "0" is registered as character size.

[0039]It returns to explanation of drawing 4 and processing which analyzes the standard of the character string attribute performing setting out of the standard of an index attribute, i.e., highlighting, can consider that it is set up, and sets up the standard is performed by the character emphasis analyzing parts 124 in the character string stored in the HTML page filter 122 in S106. The details of this processing are mentioned later.

[0040]The initial value 1 is substituted for the variable n used as a filter pointer which is a pointer for specifying one [at a time] each line (record) of the HTML filter table 122 in order by the character string analyzing parts 125 S107. It is distinguished by the character string analyzing parts 125 whether the data in which the classification of the character string shown in the line specified by the filter pointer n mentioned above in S108 is shown is "STRING", if the result of this distinction becomes in Yes, processing will progress to S109, and if No becomes, processing will progress to S115.

[0041]In S109, processing which sets an index attribute as the character string shown in the line specified by the filter pointer n based on the standard set up by processing of S106 is performed by the character string analyzing parts 125. The details of this processing are also mentioned later. Then, in [processing which starts a word from the character string shown in the line specified by the filter pointer n in S110 is performed by the character string analyzing parts 125, and] S111, Processing from which the part of speech extracts the word which is a noun is continuously performed by the character string analyzing parts 125 from the started word.

[0042]A well-known method is adopted as processing of logging of the word in S110. The method made into the word which used what is called a morphological analysis, for example, acquired the canonical form of that word for the part of speech and conjugated form of the word which were started as a method of this common knowledge using various kinds of dictionaries, and started the word of that canonical form from the character string, There are what is called an N gram method etc. that start the word of length N mechanically in order shifting logging of a character string of one character at a time from the head of the character string.

[0043]In S112, it is distinguished by the character string analyzing parts 125 whether the word extracted by the processing of S111 mentioned above existed, if this decision result becomes in Yes, processing will progress to S113, and if No becomes, processing will progress to S115. The position information which is the page which made the title the word extracted by the processing of S111 mentioned above in S113, and in which the word was contained, The index which matched with the word of the title the abstract of the text indicated to the page and the character attribute set up by processing of S109 to the character string in which the word was contained is generated by the index registering part 126, The index generated by processing of Scontinuing 114 is registered into the index file 310.

[0044]In S115, directions of the filter pointer n mentioned above by the character string analyzing parts 125 are advanced only 1. It is distinguished by the character string analyzing parts 125 whether in S116, the line specified by the present numerical value of the filter pointer n has exceeded the last line that exists in the HTML filter table 122, If this discriminated result becomes in Yes, processing will progress to S117, and if No becomes, the processing which processing returned and mentioned above to S108 will be repeated.

[0045]In S117, the directions of the page pointers m mentioned above by the Web page analyzing parts 121 are advanced only 1. It is distinguished by the Web page analyzing parts 121 whether in S118, the page specified by the present numerical value of the page pointers m has exceeded the last page of Web page 20 collected by the Web page collection Management Department 110, If the result of this distinction becomes in Yes, this index production processing will be completed. On the other hand, if the result of this discrimination processing becomes in No, the processing which processing returned and mentioned above to S104 will be repeated.

[0046]Processing to the above is index production processing. Next, the details of the standard setting processing of an index attribute performed by the character emphasis analyzing parts 124 as processing of S106 in the index production processing mentioned above are explained. Drawing 6 is a flow chart which shows the contents of processing of the standard setting processing of an index attribute.

[0047]First, in S121, the "classification" in the HTML filter table 122 is computed for every character size [in / in the sum total of the number of characters of the character string of a line which is "STRING" / a "character string attribute"]. Next, in S122, the rate of the number of characters of each character size to the number of characters of all the character strings shown, the incidence 122, i.e., the HTML filter table, of each character size, is computed.

[0048]In S123, the standard incidence S about the character size set up beforehand is acquired. In S124, it is distinguished in S125 to which total addition is carried out and the incidence computed by processing of S122 follows descending of character size whether the cumulative value exceeded the reference value S. And when this discriminated result is set to Yes, processing progresses to S126. On the other hand, while this discriminated result is No, processing of S124 is repeated.

[0049]In S126, when the result of the discrimination processing of a front step is set to Yes, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that is set up as the reference letter size Esize. In the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than this reference letter size Esize is set is a character string which is performing highlighting in Web page 20.

[0050]In S127, the "classification" in the HTML filter table 122 is computed for every character color [in / in the sum total of the number of characters of all the character strings shown in the line which is "STRING" / a "character string attribute"]. In S128, the rate of the number of characters of each character color to the number of characters of all the character strings shown, the incidence 122, i.e., the HTML filter table, of each character color, is computed.

[0051]In S129, the standard incidence C about the character color set up beforehand is acquired. In S130, it is distinguished whether the character color Cn whose incidence is more than the standard incidence C exists, and only when this discriminated result is Yes, in S131, this character color Cn is set up as the reference color Ecolor. In the standard setting processing of the index attribute mentioned later, it is considered that this reference color Ecolor is a character string to which the character string to which the attribute of a different character color is set is performing highlighting in Web page 20.

[0052]After finishing processing of S130 and S131, the standard setting processing of this index attribute is completed, and processing returns to drawing 4 mentioned above. Processing to the above is the standard setting processing of an index attribute. Next, the standard setting processing of the index attribute mentioned above is further explained using the example of drawing 8.

[0053]Drawing 8 (A) is a table in which showing the incidence of the character contained in the becoming Web page, and showing the incidence of each character size from which ** was obtained by processing to S122, and a table showing the incidence of each character color from which ** was obtained by processing to S128 document 1.

[0054]Now, suppose that the standard incidence S about the character size acquired by processing of S123 was 10%. ** It set, and the sum total of the incidence about the thing more than "5" is +5% [3% of] = 8%, and character size is less than the standard incidence S which mentioned this value above. On the other hand, it is over the standard incidence S in which the sum total of the incidence about the thing more than "4" is +70% [3% of +5% of] = 78%, and character size mentioned this value above. Therefore, in the discrimination processing of S125, when character size computes the sum total of the incidence about the thing more than "4" in processing of S124, the result serves as Yes. And in S126 performed at this time, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that, i.e., "5", is set up as the reference letter size Esize. Therefore, in the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than "5" is set, i.e., the character string by which character size is set as "5" or "6", is a character string which is performing highlighting in Web page 20.

[0055]Next, considering it as the thing of the maximum of the incidence about the character color of the character contained in the document 1 about the standard incidence C about the character color acquired by processing of S129 assumes that it was set up beforehand. ** Set, the character color of the maximum [incidence] is "black", and the incidence is 90%. Therefore, by processing of S129, since this incidence is set up as the standard incidence C, by processing of continuing S130 and S131, the black chisel whose incidence is not less than 90% is set up as the reference color Ecolor at the character color whose incidence is more than the standard incidence C, i.e., here. Therefore, in the standard setting processing of the index attribute mentioned later, it is considered that black is a character string to which the character string to which the attribute of a different character color is set, i.e., the character string whose character color is "red" or "blue" here, is performing highlighting in Web page 20.

[0056]Drawing 8 (B) is a table in which showing the incidence of the character contained in the becoming Web page, and showing the incidence of each character size from which ** was obtained by processing to S122, and a table showing the incidence of each character color from which ** was obtained by processing to S128 document 2.

[0057]Now, suppose that the standard incidence S about the character size acquired by processing of S123 was 10% like the Web page of the document 1. ** It set, and the sum total of the incidence about the thing more than "4" is 1%, and character size is less than the standard incidence S which mentioned this value above. On the other hand, it is over the standard incidence S in which the sum total of the incidence about the thing more than "3" is +99% [per %] = 100%, and character size mentioned this value above. Therefore, in the discrimination processing of S125, when character size computes the sum total of the incidence about the thing more than "3" in processing of S124, the result serves as Yes. And in S126 performed at this time, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that, i.e., "4", is set up as the reference letter size Esize. Therefore, in the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than "4" is set, i.e., the character string by which character size is set as "5", is a character string which is performing highlighting in Web page 20.

[0058]Next, suppose that the standard incidence C about the character color acquired by processing of S129 was 10%. ** It sets and the character color whose incidence is more than the standard incidence C, i.e., the red whose incidence is not less than 10% here, and blue are set up as the reference color Ecolor in processing of S130 and S131. Therefore, in the standard setting processing of the index attribute mentioned later, it is considered that it is the character string to which the attribute of a different character color from red or blue both is set, i.e., the character string to which the character string whose character color is "black" here is performing highlighting in Web page 20.

[0059]Next, the details of the setting processing of an index attribute performed by the character string analyzing parts 125 as processing of S109 in the index production processing mentioned above are explained. First, in S141, the character string attribute shown in the line specified by the current value of the filter pointer n in the HTML filter table 122 mentioned above is acquired.

[0060]In S142, the character size in the character string attribute acquired by processing of the front step, In S143 it is distinguished whether it is more than the reference letter size Esize set up by the standard setting processing of the index attribute mentioned above, and only when the result of this distinction is Yes, The character size flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0061]In S144, the character color in the character string attribute acquired by processing of S141, In S145 only when the result of this distinction is Yes, namely, only when it is distinguished whether it differs from the reference letter color Ecolor set up by the standard setting processing of the index attribute mentioned above, and character colors differ, The character color flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0062]In S147 it is distinguished whether the flag "1" is stored in the column of the "bold letter" in the character string attribute acquired by processing of S141 in S146, and only when the result of this distinction is Yes, The bold letter flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0063]After finishing processing of S146 and S147, the setting processing of this index attribute is completed and processing returns to drawing 4 mentioned above. Processing to the above is the setting processing of an index attribute. The data structure of the index file 310 generated by performing word index production processing of drawing 4 explained by the above to Web page 20 illustrated to drawing 3 is shown in drawing 9. In the figure, although the logical position on the Internet 4 to which Web page 20 shown in drawing 3 is opened is shown as "the document 1", of course, it does not matter even if URL (Uniform Resource Locator) shows this position information, for example.

[0064]If the line of the entry "easy" in drawing 9 is made into an example and the index file 310 is explained, The "easy" word is included in Web page 20 currently exhibited in the position "the document 1" [on the Internet 4] Becoming, and it is shown in Web page 20 that all of highlighting according [this word] to character size, highlighting by a character color, and highlighting by a bold letter are made. highlighting mentioned above about the "easy" word when referring to drawing 3 -- ******** -- things are shown.

[0065]Next, an example is shown and explained about the details of processing of the information retrieval performed at the information retrieval Management Department 200 which the information retrieval site 1 has. Drawing 10 is a flow chart which shows the contents of processing of the retrieval processing performed by the information management retrieval part 200. First, the contents of processing of retrieval processing are explained along with the figures.

[0066]In S201, only when it is distinguished by the information retrieval section 210 whether the search formula in which the word which are the demand of information retrieval sent from the browser 30 and an object of that search is shown was received and this discriminated result is set to Yes, processing progresses to S202. In the information retrieval section 210, if a search formula is sent, the search formula is stored in the search formula storage 211.

[0067]In S202, the sent search formula is analyzed by the information retrieval section 210, and the word which is a retrieval object is started from the search formula. Search of the entry of the index file 310 which used the search word as the key is performed by the information retrieval section 210 in the turn that the search word was started S203.

[0068]In S204, as a result of search at a front step, it is distinguished by the information retrieval section 210 whether the entry which is in agreement with a search word was discovered, if the result of this distinction becomes in Yes, processing will progress to S205, and if No becomes, processing will progress to S207. In S206 which all of position information, an abstract, and an attribute flag are acquired from the retrieval record in which the entry which is in agreement with a search word was contained by the information retrieval section 210, and continues by it in S205, The record which consists of an entry which is in agreement with this search word, and position information, an abstract and an attribute flag is stored in the search-results file 320 by the information retrieval section 210.

[0069]It is distinguished [which was mentioned above about all the search words started in S207 by the processing of S202 mentioned above] by the information retrieval section 210 whether processing of search of S203 was performed, If the result of this distinction becomes in Yes, processing will progress to S208, if the result of this distinction becomes on the other hand in No, processing will return to S203 and processing mentioned above about the search word in which processing of search is not yet performed will be performed.

[0070]The position information applicable to all the search words started by the processing of S202 mentioned above among the position information stored in the search-results file 320 in S208 here, That is, the position information included common to all the records of the search-results file 320 is extracted from the search-results file 320 by the search-results Management Department 220 with an abstract.

[0071]In S209. [whether processing at a front step was able to extract position information, and] That is, it is distinguished by the search-results Management Department 220 whether the position information included common to all the records of the search-results file 320 existed, if the result of this distinction becomes in Yes, processing will progress to S210, and if No becomes, processing will progress to S213.

[0072]In S210, the position information extracted by the processing of S208 mentioned above and the abstract which is matched and is stored in the position information in the search-results file 320 are stored in the search-results list file 330 by the search-results Management Department 220.

[0073]The number of the attribute flag which is matched and is stored in the position information extracted by the processing of S208 mentioned above in the search-results file 320 in S211 is calculated by the search-results Management Department 220 for every position information of the, This counting result is stored in the search-results list file 330 as an attribute point size.

[0074]In S212, the search-results list file 330 is sorted by the search-results Management Department 220 so that it may become descending of the enumerated data of the attribute point calculated by the front step. In S214 which the HTML file which expresses the contents of the search-results list after sorting by a Web page in S213 based on the search-results list file 330 is created by the HTML preparing part 410, and follows, The created HTML file addresses to the browser 30 which is the transmitting origin of the search formula mentioned above by the WWW server part 400, and is sent out to it, and this retrieval processing is completed.

[0075]Processing to the above is retrieval processing. Next, the case where what was shown in drawing 9 as the index file 310 is stored in the information database Management Department 300 is made into an example, and this retrieval processing is explained. First, if a search formula is sent from the browser 30, the result of distinction of S201 will serve as Yes, and logging of a search word will be performed in S202 continuing. Here, each word of the "hamburger" and the "tomato" should be started as a search word as a result of processing of S202.

[0076]If logging of a search word is completed, processing will progress to S203, first, search of a search word "hamburger" is performed about the entry of the index file 310, and the record about the entry "hamburger" in drawing 9 is discovered. Therefore, the result of the discrimination processing of S204 serves as Yes, and processing follows it to S205.

[0077]In S205, all of position information, an abstract, and an attribute flag are acquired from the discovered record, and the record which consists of the position information, an abstract, and an attribute flag in S206 continuing is stored in the search-results file 320. Then, although discrimination processing in S207 is performed, since processing of search of S203 is not yet performed about the inside "tomato" of the search word started by the processing of S202 mentioned above, the result of the discrimination processing of S207 serves as No, and processing returns to S203.

[0078]Henceforth, the same processing as the search word "hamburger" mentioned above about the search word "tomato" is performed, The record which the record about an entry "tomato" is discovered from the index file 310 shown in drawing 9, and consists of the position information, character string, and the link flag and the search word "tomato" in the record is stored in the search-results file 320.

[0079]The contents of the search-results file 320 generated by processing to the above are shown in drawing 11. After the search-results file 320 shown in this drawing 11 is generated, the result of the discrimination processing of S207 serves as No, and processing progresses to S208.

[0080]Extraction of the position information included common to all the records of the search-results file 320 in S208 is performed, as a result -- three, "the document 1", the "document 2", and the "document 3", are extracted here as position information included common to a "hamburger" and both the records of a "tomato" -- ** -- it carries out. Therefore, the result of the discrimination processing of Scontinuing 209 serves as Yes, and processing progresses to S210.

[0081]Three position information extracted in S210, "the document 1", the "document 2", and the "document 3", In S211 which the character string which is matched and is stored in the position information in the search-results file 320 is stored in the search-results list file 330, and continues, The number of the attribute flag which is matched and is stored in each of the extracted position information "document 1", the "document 2", and the "document 3" is calculated, respectively, and the counting result is stored in the search-results list file 330 as an attribute point size.

[0082]Drawing 12 is explained here. The figure shows the contents of the search-results list file 330, and what is shown in the figure (a) is created as the search-results list file 330 by the processing to S211 mentioned above. In the search-results file 320 shown in drawing 11, since a total of six attribute flags about "the document 1" are stored, the attribute point size about the position information "document 1" in the search-results list file 330 shown in drawing 12 (a) is set to "6."

[0083]From the search-results file 320 same also about the attribute point of the "document 2" and the "document 3" and shown in drawing 11. The attribute point size about the "document 2" in the search-results list file 330 shown in drawing 12 (a) is set to "0", and the attribute point size about "the document 3" is set to "3."

[0084]In S212 continuing, creation of the search-results list file 330 which shows drawing 12 (a) the contents by processing to S211 mentioned above will perform sorting of the search-results list file 330 so that it may become descending of an attribute point size value. The result to which sorting based on an attribute point size was carried out to the search-results list file 330 of drawing 12 (a) is shown in drawing 12 (b), and the turn of each line is rearranged in order of the "document 1" with a high attribute point, the "document 3", and the "document 2."

[0085]Then, in S213, the HTML file which the HTML file which expresses the contents of the search-results list file 330 to which sorting was performed like drawing 12 (b) by a Web page was created, and was created in S214 continuing is sent out, and this retrieval processing is completed.

[0086]The example of a screen of the Web page which shows the result of information retrieval displayed when the created HTML file is perused by the browser 30 is shown in drawing 13. . Are preferentially arranged from what it is the search results about the word of a "hamburger" and a "tomato", and is expected to acquire the information that importance is high in the screen shown in the figure. The hyperlink to Web page 20 respectively shown by that position information is embedded at the position information on "the document 1", the "document 3", and the "document 2", and facilities are given to the user of these search results.

[0087]In the analysis of the HTML sauce about Web page 20 in the embodiment described by the above, although it has judged whether the character string is emphasized in Web page 20 based on description of the <FONT> tag and the <B> tag, Based on description of other tags, it may be made to perform this judgment. As an example of a tag employable as the judgment of this emphasis, <I> on which a character string is displayed with an italic character The <U> tag which gives an underline to a tag or a character string, Or at a standard browser, there are a <STRONG> tag etc. which can enable it to make the character string able to pronounce strongly by the voice browser which reads out with a sound the text indicated to the Web page of that on which a character string is only only displayed in a bold letter character. Based on the <FONT> tag in which the FACE attribute for specifying the kind of font used for the display of a character string is specified, It may be made for the character string as which a different font from that by which normal use is carried out to the display sake of the Web page is specified to judge with what is emphasized in the Web page.

[0088]The control program for making it carry out to the computer which has the standard composition which mentioned above the index production processing and retrieval processing which the information site 1 was performing in the embodiment of explained this invention by the above, and the same processing is created, By making the control program read into the computer, and performing it, this invention can be carried out by such computer.

[0089]It is also possible to carry out this invention by computer by making such a control program record on the recording medium which can be read by computer, making the program read from a recording medium to a computer, and performing it. The example of the possible recording medium of reading the control program made to record by computer is shown in drawing 14. The memory storage 502, such as ROM with which the computer 501 is equipped as built-in or external attachment as a recording medium, for example as shown in the figure, and a hard disk drive, or a flexible disk, Portable good signifier recording-media 503 grades, such as MO (magneto-optical disc), CD-ROM, and DVD-ROM, can be used. A recording medium may be the memory storage 506 which is connected with the computer 501 via the network 504 and with which the computer which functions as the program server 505 is provided. In this case, the transmission signal acquired by modulating a subcarrier with the data signal expressing a control program, The control program concerned can be executed now by making it transmit through the network 504 which is a transmission medium from the program server 5055, restoring to the transmission signal received by computer 501, and reproducing a control program.

[0090]

[Effect of the Invention]According to this invention, in the word which constitutes the character string contained in the document information currently released on the communication network. By matching the position information which shows the position of the document information, and the emphasis attributes which show the emphasis given to the character string, and registering with an index file. When the index file is searched based on the word showing a retrieval object, Since that with which emphasis attributes are matched can be given priority to and shown to the word matched with the position information among the position information acquired by the search, an information retrieval person can be provided with a more suitable information retrieval result to a retrieval object.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Field of the Invention]This invention relates to the art of enabling it to provide the information which agreed appropriately by the demand, to the demand of search especially about the art of retrieving information.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Description of the Prior Art]In recent years, the number of the Web pages provided by the WWW (WorldWideWeb) system on the Internet is continuing increasing explosively by the spread of the Internet. On the Internet, many search engines which provide the service which retrieves the information made into the purpose out of this huge information are established.
[0003]There are some which are called the robot type as one of the methods which collects the information on a network of a search engine. In a robot type search engine, the robot program called a spider or a crawler is started periodically, Automatic collection of the HTML (HyperText Markup Language) file expressing the Web page currently exhibited on the Internet is performed. When information retrieval is performed and the information retrieval person using a search engine gives a closely related keyword to the target information at a search engine site, Processing which extracts that in which the keyword was contained from the collected file is performed, and an information retrieval person is provided with the list of Web pages which are the keyword and which are contained as search results with the information which shows the logical position on the Internet about the Web page.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Effect of the Invention]In the word which constitutes the character string contained in the document information currently released on the communication network from this invention. By matching the position information which shows the position of the document information, and the emphasis attributes which show the emphasis given to the character string, and registering with an index file. When the index file is searched based on the word showing a retrieval object, that by which emphasis attributes are matched with the word matched with the position information among the position information acquired by the search can be given priority to and shown.
Therefore, an information retrieval person can be provided with a more suitable information retrieval result to a retrieval object.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Problem(s) to be Solved by the Invention]Generally, since the robot type search engine is performing [no] automatically processings of a to [from collection of information / offer of search results] by computer and operation of the information by judgment of human being intervenes there, The arrangement about the quality of the genre to which the collected information belongs, or its information is not made. Therefore, if search by coincidence of a mere keyword was performed on the occasion of search of information, a Web page including important information is buried in search results, or. Or there were not few cases where it will be mostly contained in search results only in about the Web page in which what is called a search noise, i.e., the low information on usefulness, is contained.
[0005]Making more suitable the result of the information retrieval which a search engine provides in view of the above problem to an information retrieval person's retrieval object is the issue which this invention tends to solve.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Means for Solving the Problem]A word contained in document information to which this invention is opened on a communication network, An index file which matches position information which shows a document information position in which it is the information which shows a logical position on this communication network, and information containing this word exists is prepared, It is premised on a system or a method of showing position information corresponding to a word which searches this index file based on a word showing an object of search, and expresses this retrieval object.
[0007]And an information retrieval system which is one of the modes of this invention, An emphasis-attributes acquisition means which acquires emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis, In a word extracted by extraction means to extract a word from said character string, and said extraction means. A registration means to match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word, and to register with said index file, A search means to acquire position information which searches said index file based on a word showing said retrieval object, and is matched with this word from this index file, A presenting means which gives priority to that by which said emphasis attributes are matched with a word which this search means made an object of search of this position information in said index file among position information acquired by said search means, and presents this position information, SUBJECT mentioned above is solved by constituting so that it may ****.
[0008]An attribute given to a character string contained in said document information here is an attribute which shows color of a character used in order to display an attribute which shows a size of a character used when displaying this character string, for example, or this character string. It is possible that document information containing a word emphasis is instructed to be has a high possibility that information that importance is high is included about the word. Therefore, when two or more document information corresponding to a word which shows a search condition is opened to a communication network according to composition mentioned

above, Since the word is emphasized, and priority is given to position information about document information considered that importance is high, it makes and it is shown, a result of information retrieval will become more suitable to an information retrieval person's retrieval object.

[0009]In an information retrieval system concerning this invention mentioned above, said emphasis-attributes acquisition means, A frequency-of-occurrence calculating means which computes the frequency of occurrence in this document information about an attribute given to a character string contained in said document information for this every attribute, An emphasis-attributes setting-out means to set up a standard which distinguishes whether said attributes are said emphasis attributes based on the frequency of occurrence for this every attribute, It may constitute so that it may have an emphasis-attributes discriminating means which distinguishes whether attributes given to a character string contained in said document information are said emphasis attributes based on said standard.

[0010]Since it can judge now that a character string to which a unique attribute is given in document information is emphasized in that document information according to this composition, this document information can be registered into an index file for a word contained in this character string as what has high importance.

[0011]In an information retrieval system concerning this invention mentioned above, said emphasis-attributes acquisition means, When making into a bold letter a character used in order that an attribute given to a character string contained in said document information may display this character string is shown, it may be made to consider that these attributes are said emphasis attributes.

[0012]According to this composition, in order to show emphasis of a character string in document information, the attribute of a purport that a display by a bold letter currently performed widely is performed can be promptly judged with emphasis attributes. In an information retrieval system concerning this invention mentioned above, said presenting means, Priority is given to a thing which has many number of said emphasis attributes matched with a word which this search means made an object of search of this position information in said search file among position information acquired by said search means, and it may be made to show this position information.

[0013]According to this composition, for a word by which a thing which has many number of emphasis attributes given to a certain character string is contained in this character string, this document information can be registered now into an index file as what has higher importance, and a result of information retrieval will become still more suitable to an information retrieval person's retrieval object.

[0014]An information retrieval method which is one of the another modes of this invention, Emphasis attributes which are attributes given to a character string contained in said document information, and show emphasis are acquired, Match said position information about this word, and said emphasis attributes given to a character string of extraction origin of this word with a word which extracted a word from said character string and was extracted from said character string, and it registers with said index file, Position information which searches said index file based on a word showing said retrieval object, and is matched with this word is acquired from this index file, The same operation and effect as an information retrieval system concerning this invention mentioned above are acquired by giving priority to that by which said emphasis attributes are matched with a word made into an object of search of this position information in said search file among position information acquired by said search, and showing this position information.

[0015]SUBJECT mentioned above by making a computer execute the program also in a program for making processing which consists of the same procedure as an information retrieval method concerning this invention mentioned above perform to a computer is solvable.

[0016]

[Embodiment of the Invention]Hereafter, an embodiment of the invention is described based on a drawing. Drawing 1 is a figure in which the information retrieval site which carries out this invention shows the entire configuration of the communication network which provides an information search service.

[0017]In drawing 1, the Internet 4 which is all a communication network is accessed, and data can be delivered [the information retrieval site 1 the offer-of-information site 2a, 2b, 2c, 2d, and the user terminals 3a and 3b] and received mutually. The information retrieval site 1 is a WWW server system which provides a robot search type information search service for the user terminals 3a and 3b, is provided with the Research and Data Processing Department 100, the information retrieval Management Department 200, the information database Management Department 300, and the WWW server Management Department 400, and is constituted.

[0018]The Research and Data Processing Department 100 performs automatic collection of the information currently released on the Internet 4, and accumulates the collected information in the information database Management Department 300. The information retrieval Management Department 200 retrieves the information accumulated in the information database Management Department 300 according to the demand of the information retrieval sent via the Internet 4, and returns the result of the search to a requiring agency.

[0019]At the information database Management Department 300, accumulation of the information collected by the Research and Data Processing Department 100 and search of the information by the information retrieval Management Department 200 are performed. The processing by which the WWW server part 400 transmits the collected information which is sent via the Internet 4 to the Research and Data Processing Department 100, Processing which transmits the demand of the information retrieval sent via the Internet 4 to the information retrieval Management Department 200, and processing of sending out of a Web page in which the information which shows the result of the information retrieval sent from the information retrieval Management Department 200 is expressed are performed.

[0020]The offer-of-information site 2a, 2b, and 2c and 2d are WWW server systems which exhibit Web pages 20a, 20b, 20c, and 20d on the Internet 4, respectively. Although four offer-of-information sites are shown in drawing 1, the number of the offer-of-information sites connected to the Internet 4 may be arbitrary.

[0021]The user terminals 3a and 3b, respectively The offer-of-information site 2a, 2b, 2c, And it is a computer which can perform the browsers 30a and 30b which are the software which peruses the Web page provided from 2 d or the information retrieval site 1, and is operated by the information retrieval person who requests search of the information currently released on the Internet 4 to the information retrieval site 1. Although two users are shown in drawing 1, the number of the user terminals connected to the Internet 4 may also be arbitrary.

[0022]These information retrieval sites 1, the offer-of-information site 2a, 2b, 2c and 2d, and the user terminals 3a and 3b, The computer by which all have standard composition, i.e., CPU which controls each component by executing a control program, The storage parts store used as the work area at the time of the memory and CPU of a control program which consist of a ROM, RAM, a magnetic storage device, etc., and make CPU control each component executing a control program, or a storage area of various data, It can also constitute using a computer provided with the input part from which various kinds of data corresponding to operation by a user is acquired, the outputting part which show a display etc. various kinds of data and of which a user is notified, and the I/F part which provides the interface function for connecting with a network.

[0023]Next, drawing 2 is explained. The figure shows the detailed composition of the information retrieval site 1 in drawing 1 which carries out this invention. As shown in drawing 2, the Research and Data Processing Department 100 has the Web page collection Management Department 110 and the index production part 120, and is constituted, The information management retrieval part 200 is provided with the information retrieval section 210 and the search-results Management Department 220, and is constituted, The database manager 300 has the index file 310, the search-results file 320, and the search-results list file 330, and is constituted, and WWW server 400 is provided with the HTML preparing part 410, and is constituted.

[0024]The Web page collection Management Department 110 performs periodically automatic collection of Web page 20 currently exhibited on the Internet 4. The position information on Web page 20 by which the index build Management Department 120 was collected by the Web page collection Management Department 110, That is, the record used as the index which can lengthen the position information which shows the logical position on the Internet 4 in which Web page 20 exists is created, and it registers with the index file 310. The index build Management Department 120 has the Web page analyzing parts 121, the word extraction Management Department 123, and the index registering part 126, and is constituted.

[0025]The Web page analyzing parts 121 create the HTML filter table 122 which makes the unit of a record each HTML tag described by the text of the HTML file which analyzes Web page 20 and is expressing Web page 20. processing which analyzes the conditions of the character style which is emphasized when they are displayed as a screen of Web page 20 in the character string shown in the HTML filter table 122, and which can be been [character style / it] rich and made being performed by the character emphasis analyzing parts 124, and at the word extraction Management Department 123, Analysis of the character string shown in the HTML filter table 122 is conducted by the character string analyzing parts 125, and a word is extracted from the character string.

[0026]The logical position information [the index registering part 126 makes an entry the word extracted by the character string analyzing parts 125, and / entry / the] on the Internet 4 about Web page 20, The index record in which the abstract of Web page 20 in which the word was contained, and the form set as the word by Web page 20 matched the attribute flag which shows that it agrees on the conditions of the character style obtained in the analysis in the character emphasis analyzing parts 124 is registered into the index file 310.

[0027]The information retrieval section 210 acquires the demand of the information retrieval sent from the user terminal by control of the browser 30 currently performed with one of the user terminals connected to the Internet 4 from the WWW server part 400, The search formula showing the conditions of the information retrieval is taken out from the demand, and it stores in the search formula storage 211. And the word (keyword) which searches the index file 310 and is shown in the search formula acquires the index record used as a title, and stores in the search-results file 320.

[0028]If search by the information retrieval section 210 is completed, the search-results Management Department 220, The position information and the abstract which are shown in the index data stored in the search-results file 320, and the total number of the attribute flag attached with the index record corresponding to the position information are stored in the search-results list file 330. And the position information stored in the search-results list file 330 is sorted according to the total number.

[0029]The HTML preparing part 410 creates the HTML file expressing the Web page which receives the search-results list which consists of sorted position information which is stored in the search-results list file 330 and as which the search-results list is expressed. The created HTML file is addressed to the user terminal in which the browser 30 is performed, and is sent out to the Internet 4 by the WWW server part 400.

[0030]Next, the details of processing of the collection of a Web page performed in the Research and Data Processing Department 100 which the information retrieval site 1 has, and generation of an index are explained. Drawing 3 shows the example of Web page 20 which is opened to the Internet 4 and collected by the information retrieval site 1. In the figure, the browser's 30 inspection of the HTML sauce shown in (b) will display the screen shown in the figure (a).

[0031]Drawing 4 is explained here. The figure is a flow chart which shows the contents of processing of the index production processing performed in the Research and Data

Processing Department 100. By performing this processing, collection of a Web page and generation of an index are performed in the Research and Data Processing Department 100. First, in S101, only when it was distinguished at the Web page collection Management Department 110 whether the present date is the collection designated date of Web page 20 specified beforehand, this decision result is set to Yes and the present becomes that designated date, processing progresses to S102. Although the method of specification of this date is arbitrary, specification called the monthly final day of month end, etc. is performed, for example.

[0032]In S102, processing of a round and collection of Web page 20 currently exhibited on the Internet 4 is performed by the Web page collection Management Department 110. The technique of this round and collection should just use as it is what is performed from the former by the well-known robot type search engine.

[0033]In the Web page analyzing parts 121, the initial value 1 is substituted for the variable m used as page pointers which are pointers for pointing at a time to 1 page of Web pages 20 of a large number collected at the front step S103. The structure of the page to which it points by the current value of the page pointers m in Web page 20 collected by processing of S102 in S104 is analyzed by the Web page analyzing parts 121, The HTML filter table 122 is generated by the Web page analyzing parts 121 in S105 continuing.

[0034]The HTML filter table generated from the Web page shown in drawing 3 is shown in drawing 5. Analysis of the HTML sauce shown in drawing 3 (b) by the Web page analyzing parts 121 will generate the HTML filter table shown in drawing 5. When it explains further, referring to drawing 3 (b) for the contents of processing of S104, in the Web page analyzing parts 121. The text of the HTML sauce of an analytical object, i.e., the portion pinched between the start tag of <BODY> and the end tag, is made into the object of analysis, and the structure of each sentence where the <BR> tag (line feed tag) in the portion is considered as a pause of a sentence, and is contained in the text is analyzed.

[0035]If signs that the HTML filter table shown in drawing 5 from the HTML sauce shown in drawing 3 (b) is created are explained, First, let the portion pinched between the start tag of the <BODY> tag and end tag which are the description parts of the text in HTML sauce, i.e., the portion put between the <BODY> tag and the </BODY> tag, be an object of the analysis by processing of S104.

[0036]here -- first -- an analytical object -- a portion -- it can set -- the beginning -- < -- BR -- > -- a tag -- describing -- having -- **** -- a part -- up to -- a portion -- namely, -- "-- < -- FONT SIZE -- = -- " -- six -- " -- COLOR -- = -- " -- # -- FF -- 0000 -- " -- > -- < -- B -- > -- easy -- cooking -- < -- /-- B -- > -- < -- /-- FONT -- > -- " -- a portion is analyzed. <FONT SIZE="6"COLOR="#FF0000"> Here the becoming tag, the "easy dish" described by the portion pinched between the start tag of <FONT>, and the end tag -- a character string -- character size -- "6" -- considering it as a size -- and a character color -- "#FF0000" -- what is displayed as a color shown numerically is shown. The color shown for this figure is red.

[0037]Displaying the character string "the easy dish" <B> [which is described by the portion by which the becoming tag is sandwiched between the start tag of <FONT> and the end tag] Becoming in a bold letter is shown. The analysis processing by S104 conducts analysis mentioned above, and the record shown in the 1st line of drawing 5 which means the contents of this analysis result is stored in the HTML page filter 122 by processing of S105 performed next. An "easy dish" is stored in the column of a "character string" when this record is explained, "STRING" which shows that an "easy dish" is a character string is stored in the column of "classification", and the attribute which shows that "6" and a "color" are displayed for "character size" as "red" about the character string of an "easy dish" is stored in the column of a "character string attribute." By furthermore storing the flag "1" in the column of the "bold letter" in a "character string attribute" shows that displaying the character string of an "easy dish" in a bold letter was shown. And the record in which existence of the <BR> tag which only makes it the contents for "classification" to be "BR" after the is shown is stored in the 2nd line of the HTML page filter 122.

[0038]Analysis of the remaining portion in the text part of the HTML sauce shown in drawing 3 (b) is conducted similarly hereafter, and the HTML page filter 122 shown in drawing 5 in this way is generated. Although the SIZE attribute in the <FONT> tag may be shown like "+1", "-1", etc. by the relative value over the usual displayed character size, When such, the relative value is registered as character size, and about the character string by which specification of character size is not made, "0" is registered as character size.

[0039]It returns to explanation of drawing 4 and processing which analyzes the standard of the character string attribute performing setting out of the standard of an index attribute, i.e., highlighting, can consider that it is set up, and sets up the standard is performed by the character emphasis analyzing parts 124 in the character string stored in the HTML page filter 122 in S106. The details of this processing are mentioned later.

[0040]The initial value 1 is substituted for the variable n used as a filter pointer which is a pointer for specifying one [at a time] each line (record) of the HTML filter table 122 in order by the character string analyzing parts 125 S107. It is distinguished by the character string analyzing parts 125 whether the data in which the classification of the character string shown in the line specified by the filter pointer n mentioned above in S108 is shown is "STRING", if the result of this distinction becomes in Yes, processing will progress to S109, and if No becomes, processing will progress to S115.

[0041]In S109, processing which sets an index attribute as the character string shown in the line specified by the filter pointer n based on the standard set up by processing of S106 is performed by the character string analyzing parts 125. The details of this processing are also mentioned later. Then, in [processing which starts a word from the character string shown in the line specified by the filter pointer n in S110 is performed by the character string analyzing parts 125, and] S111, Processing from which the part of speech extracts the word which is a noun is continuously performed by the character string analyzing parts 125 from the started word.

[0042]A well-known method is adopted as processing of logging of the word in S110. The method made into the word which used what is called a morphological analysis, for example, acquired the canonical form of that word for the part of speech and conjugated form of the word which were started as a method of this common knowledge using various kinds of dictionaries, and started the word of that canonical form from the character string, There are what is called an N gram method etc. that start the word of length N mechanically in order shifting logging of a character string of one character at a time from the head of the character string.

[0043]In S112, it is distinguished by the character string analyzing parts 125 whether the word extracted by the processing of S111 mentioned above existed, if this decision result becomes in Yes, processing will progress to S113, and if No becomes, processing will progress to S115. The position information which is the page which made the title the word extracted by the processing of S111 mentioned above in S113, and in which the word was contained, The index which matched with the word of the title the abstract of the text indicated to the page and the character attribute set up by processing of S109 to the character string in which the word was contained is generated by the index registering part 126, The index generated by processing of Scontinuing 114 is registered into the index file 310.

[0044]In S115, directions of the filter pointer n mentioned above by the character string analyzing parts 125 are advanced only 1. It is distinguished by the character string analyzing parts 125 whether in S116, the line specified by the present numerical value of the filter pointer n has exceeded the last line that exists in the HTML filter table 122, If this discriminated result becomes in Yes, processing will progress to S117, and if No becomes, the processing which processing returned and mentioned above to S108 will be repeated.

[0045]In S117, the directions of the page pointers m mentioned above by the Web page analyzing parts 121 are advanced only 1. It is distinguished by the Web page analyzing parts 121 whether in S118, the page specified by the present numerical value of the page pointers m has exceeded the last page of Web page 20 collected by the Web page collection Management Department 110, If the result of this distinction becomes in Yes, this index production processing will be completed. On the other hand, if the result of this discrimination processing becomes in No, the processing which processing returned and mentioned above to S104 will be repeated.

[0046]Processing to the above is index production processing. Next, the details of the standard setting processing of an index attribute performed by the character emphasis analyzing parts 124 as processing of S106 in the index production processing mentioned above are explained. Drawing 6 is a flow chart which shows the contents of processing of the standard setting processing of an index attribute.

[0047]First, in S121, the "classification" in the HTML filter table 122 is computed for every character size [in / in the sum total of the number of characters of the character string of a line which is "STRING" / a "character string attribute"]. Next, in S122, the rate of the number of characters of each character size to the number of characters of all the character strings shown, the incidence 122, i.e., the HTML filter table, of each character size, is computed.

[0048]In S123, the standard incidence S about the character size set up beforehand is acquired. In S124, it is distinguished in S125 to which total addition is carried out and the incidence computed by processing of S122 follows descending of character size whether the cumulative value exceeded the reference value S. And when this discriminated result is set to Yes, processing progresses to S126. On the other hand, while this discriminated result is No, processing of S124 is repeated.

[0049]In S126, when the result of the discrimination processing of a front step is set to Yes, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that is set up as the reference letter size Esize. In the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than this reference letter size Esize is set is a character string which is performing highlighting in Web page 20.

[0050]In S127, the "classification" in the HTML filter table 122 is computed for every character color [in / in the sum total of the number of characters of all the character strings shown in the line which is "STRING" / a "character string attribute"]. In S128, the rate of the number of characters of each character color to the number of characters of all the character strings shown, the incidence 122, i.e., the HTML filter table, of each character color, is computed.

[0051]In S129, the standard incidence C about the character color set up beforehand is acquired. In S130, it is distinguished whether the character color Cn whose incidence is more than the standard incidence C exists, and only when this discriminated result is Yes, in S131, this character color Cn is set up as the reference color Ecolor. In the standard setting processing of the index attribute mentioned later, it is considered that this reference color Ecolor is a character string to which the character string to which the attribute of a different character color is set is performing highlighting in Web page 20.

[0052]After finishing processing of S130 and S131, the standard setting processing of this index attribute is completed, and processing returns to drawing 4 mentioned above. Processing to the above is the standard setting processing of an index attribute. Next, the standard setting processing of the index attribute mentioned above is further explained using the example of drawing 8.

[0053]Drawing 8 (A) is a table in which showing the incidence of the character contained in the becoming Web page, and showing the incidence of each character size from which ** was obtained by processing to S122, and a table showing the incidence of each character color from which ** was obtained by processing to S128 document 1.

[0054]Now, suppose that the standard incidence S about the character size acquired by processing of S123 was 10%. ** It set, and the sum total of the incidence about the thing more than "5" is +5% [3% of] = 8%, and character size is less than the standard incidence S which mentioned this value above. On the other hand, it is over the standard incidence S in which the sum total of the incidence about the thing more than "4" is +70% [3% of +5% of] = 78%, and character size mentioned this value above. Therefore, in the discrimination processing of S125, when character size computes the sum total of the incidence about the thing more than "4" in processing of S124, the result serves as Yes. And in S126 performed at this time, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that, i.e., "5", is set up as the reference letter size Esize. Therefore, in the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than "5" is set, i.e., the character string by which character size is set as "5" or "6", is a character string which is performing highlighting in Web page 20.

[0055]Next, considering it as the thing of the maximum of the incidence about the character color of the character contained in the document 1 about the standard incidence C about the character color acquired by processing of S129 assumes that it was set up beforehand. ** Set, the character color of the maximum [incidence] is "black", and the incidence is 90%. Therefore, by processing of S129, since this incidence is set up as the standard incidence C, by processing of continuing S130 and S131, the black chisel whose incidence is not less than 90% is set up as the reference color Ecolor at the character color whose incidence is more than the standard incidence C, i.e., here. Therefore, in the standard setting processing of the index attribute mentioned later, it is considered that black is a character string to which the character string to which the attribute of a different character color is set, i.e., the character string whose character color is "red" or "blue" here, is performing highlighting in Web page 20.

[0056]Drawing 8 (B) is a table in which showing the incidence of the character contained in the becoming Web page, and showing the incidence of each character size from which ** was obtained by processing to S122, and a table showing the incidence of each character color from which ** was obtained by processing to S128 document 2.

[0057]Now, suppose that the standard incidence S about the character size acquired by processing of S123 was 10% like the Web page of the document 1. ** It set, and the sum total of the incidence about the thing more than "4" is 1%, and character size is less than the standard incidence S which mentioned this value above. On the other hand, it is over the standard incidence S in which the sum total of the incidence about the thing more than "3" is +99% [per %] = 100%, and character size mentioned this value above. Therefore, in the discrimination processing of S125, when character size computes the sum total of the incidence about the thing more than "3" in processing of S124, the result serves as Yes. And in S126 performed at this time, the larger character size by one than the character size corresponding to the incidence added by the processing of S124 in front of that, i.e., "4", is set up as the reference letter size Esize. Therefore, in the standard setting processing of the index attribute mentioned later, it considers that the character string to which the attribute of the character size more than "4" is set, i.e., the character string by which character size is set as

"5", is a character string which is performing highlighting in Web page 20.

[0058]Next, suppose that the standard incidence C about the character color acquired by processing of S129 was 10%. ** It sets and the character color whose incidence is more than the standard incidence C, i.e., the red whose incidence is not less than 10% here, and blue are set up as the reference color Ecolor in processing of S130 and S131. Therefore, in the standard setting processing of the index attribute mentioned later, it is considered that it is the character string to which the attribute of a different character color from red or blue both is set, i.e., the character string to which the character string whose character color is "black" here is performing highlighting in Web page 20.

[0059]Next, the details of the setting processing of an index attribute performed by the character string analyzing parts 125 as processing of S109 in the index production processing mentioned above are explained. First, in S141, the character string attribute shown in the line specified by the current value of the filter pointer n in the HTML filter table 122 mentioned above is acquired.

[0060]In S142, the character size in the character string attribute acquired by processing of the front step, In S143 it is distinguished whether it is more than the reference letter size Esize set up by the standard setting processing of the index attribute mentioned above, and only when the result of this distinction is Yes, The character size flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0061]In S144, the character color in the character string attribute acquired by processing of S141, In S145 only when the result of this distinction is Yes, namely, only when it is distinguished whether it differs from the reference letter color Ecolor set up by the standard setting processing of the index attribute mentioned above, and character colors differ, The character color flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0062]In S147 it is distinguished whether the flag "1" is stored in the column of the "bold letter" in the character string attribute acquired by processing of S141 in S146, and only when the result of this distinction is Yes, The bold letter flag with which the attribute given to the character string shown in the line specified by the current value of the filter pointer n mentioned above is defined as the index attribute buffer held temporarily is set to "1."

[0063]After finishing processing of S146 and S147, the setting processing of this index attribute is completed and processing returns to drawing 4 mentioned above. Processing to the above is the setting processing of an index attribute. The data structure of the index file 310 generated by performing word index production processing of drawing 4 explained by the above to Web page 20 illustrated to drawing 3 is shown in drawing 9. In the figure, although the logical position on the Internet 4 to which Web page 20 shown in drawing 3 is opened is shown as "the document 1", of course, it does not matter even if URL (Uniform Resource

Locator) shows this position information, for example.

[0064]If the line of the entry "easy" in drawing 9 is made into an example and the index file 310 is explained, The "easy" word is included in Web page 20 currently exhibited in the position "the document 1" [on the Internet 4] Becoming, and it is shown in Web page 20 that all of highlighting according [this word] to character size, highlighting by a character color, and highlighting by a bold letter are made. highlighting mentioned above about the "easy" word when referring to drawing 3 -- ******** -- things are shown.

[0065]Next, an example is shown and explained about the details of processing of the information retrieval performed at the information retrieval Management Department 200 which the information retrieval site 1 has. Drawing 10 is a flow chart which shows the contents of processing of the retrieval processing performed by the information management retrieval part 200. First, the contents of processing of retrieval processing are explained along with the figures.

[0066]In S201, only when it is distinguished by the information retrieval section 210 whether the search formula in which the word which are the demand of information retrieval sent from the browser 30 and an object of that search is shown was received and this discriminated result is set to Yes, processing progresses to S202. In the information retrieval section 210, if a search formula is sent, the search formula is stored in the search formula storage 211.

[0067]In S202, the sent search formula is analyzed by the information retrieval section 210, and the word which is a retrieval object is started from the search formula. Search of the entry of the index file 310 which used the search word as the key is performed by the information retrieval section 210 in the turn that the search word was started S203.

[0068]In S204, as a result of search at a front step, it is distinguished by the information retrieval section 210 whether the entry which is in agreement with a search word was discovered, if the result of this distinction becomes in Yes, processing will progress to S205, and if No becomes, processing will progress to S207. In S206 which all of position information, an abstract, and an attribute flag are acquired from the retrieval record in which the entry which is in agreement with a search word was contained by the information retrieval section 210, and continues by it in S205, The record which consists of an entry which is in agreement with this search word, and position information, an abstract and an attribute flag is stored in the search-results file 320 by the information retrieval section 210.

[0069]It is distinguished [which was mentioned above about all the search words started in S207 by the processing of S202 mentioned above] by the information retrieval section 210 whether processing of search of S203 was performed, If the result of this distinction becomes in Yes, processing will progress to S208, if the result of this distinction becomes on the other hand in No, processing will return to S203 and processing mentioned above about the search word in which processing of search is not yet performed will be performed.

[0070]The position information applicable to all the search words started by the processing of S202 mentioned above among the position information stored in the search-results file 320 in

S208 here, That is, the position information included common to all the records of the search-results file 320 is extracted from the search-results file 320 by the search-results Management Department 220 with an abstract.

[0071]In S209. [whether processing at a front step was able to extract position information, and] That is, it is distinguished by the search-results Management Department 220 whether the position information included common to all the records of the search-results file 320 existed, if the result of this distinction becomes in Yes, processing will progress to S210, and if No becomes, processing will progress to S213.

[0072]In S210, the position information extracted by the processing of S208 mentioned above and the abstract which is matched and is stored in the position information in the search-results file 320 are stored in the search-results list file 330 by the search-results Management Department 220.

[0073]The number of the attribute flag which is matched and is stored in the position information extracted by the processing of S208 mentioned above in the search-results file 320 in S211 is calculated by the search-results Management Department 220 for every position information of the, This counting result is stored in the search-results list file 330 as an attribute point size.

[0074]In S212, the search-results list file 330 is sorted by the search-results Management Department 220 so that it may become descending of the enumerated data of the attribute point calculated by the front step. In S214 which the HTML file which expresses the contents of the search-results list after sorting by a Web page in S213 based on the search-results list file 330 is created by the HTML preparing part 410, and follows, The created HTML file addresses to the browser 30 which is the transmitting origin of the search formula mentioned above by the WWW server part 400, and is sent out to it, and this retrieval processing is completed.

[0075]Processing to the above is retrieval processing. Next, the case where what was shown in drawing 9 as the index file 310 is stored in the information database Management Department 300 is made into an example, and this retrieval processing is explained. First, if a search formula is sent from the browser 30, the result of distinction of S201 will serve as Yes, and logging of a search word will be performed in S202 continuing. Here, each word of the "hamburger" and the "tomato" should be started as a search word as a result of processing of S202.

[0076]If logging of a search word is completed, processing will progress to S203, first, search of a search word "hamburger" is performed about the entry of the index file 310, and the record about the entry "hamburger" in drawing 9 is discovered. Therefore, the result of the discrimination processing of S204 serves as Yes, and processing follows it to S205.

[0077]In S205, all of position information, an abstract, and an attribute flag are acquired from the discovered record, and the record which consists of the position information, an abstract, and an attribute flag in S206 continuing is stored in the search-results file 320. Then, although discrimination processing in S207 is performed, since processing of search of S203 is not yet performed about the inside "tomato" of the search word started by the processing of S202 mentioned above, the result of the discrimination processing of S207 serves as No, and processing returns to S203.

[0078]Henceforth, the same processing as the search word "hamburger" mentioned above about the search word "tomato" is performed, The record which the record about an entry "tomato" is discovered from the index file 310 shown in drawing 9, and consists of the position information, character string, and the link flag and the search word "tomato" in the record is stored in the search-results file 320.

[0079]The contents of the search-results file 320 generated by processing to the above are shown in drawing 11. After the search-results file 320 shown in this drawing 11 is generated, the result of the discrimination processing of S207 serves as No, and processing progresses to S208.

[0080]Extraction of the position information included common to all the records of the search-results file 320 in S208 is performed, as a result -- three, "the document 1", the "document 2", and the "document 3", are extracted here as position information included common to a "hamburger" and both the records of a "tomato" -- ** -- it carries out. Therefore, the result of the discrimination processing of Scontinuing 209 serves as Yes, and processing progresses to S210.

[0081]Three position information extracted in S210, "the document 1", the "document 2", and the "document 3", In S211 which the character string which is matched and is stored in the position information in the search-results file 320 is stored in the search-results list file 330, and continues, The number of the attribute flag which is matched and is stored in each of the extracted position information "document 1", the "document 2", and the "document 3" is calculated, respectively, and the counting result is stored in the search-results list file 330 as an attribute point size.

[0082]Drawing 12 is explained here. The figure shows the contents of the search-results list file 330, and what is shown in the figure (a) is created as the search-results list file 330 by the processing to S211 mentioned above. In the search-results file 320 shown in drawing 11, since a total of six attribute flags about "the document 1" are stored, the attribute point size about the position information "document 1" in the search-results list file 330 shown in drawing 12 (a) is set to "6."

[0083]From the search-results file 320 same also about the attribute point of the "document 2" and the "document 3" and shown in drawing 11. The attribute point size about the "document 2" in the search-results list file 330 shown in drawing 12 (a) is set to "0", and the attribute point size about "the document 3" is set to "3."

[0084]In S212 continuing, creation of the search-results list file 330 which shows drawing 12 (a) the contents by processing to S211 mentioned above will perform sorting of the search-results list file 330 so that it may become descending of an attribute point size value. The result to which sorting based on an attribute point size was carried out to the search-results list file

330 of drawing 12 (a) is shown in drawing 12 (b), and the turn of each line is rearranged in order of the "document 1" with a high attribute point, the "document 3", and the "document 2."

[0085]Then, in S213, the HTML file which the HTML file which expresses the contents of the search-results list file 330 to which sorting was performed like drawing 12 (b) by a Web page was created, and was created in S214 continuing is sent out, and this retrieval processing is completed.

[0086]The example of a screen of the Web page which shows the result of information retrieval displayed when the created HTML file is perused by the browser 30 is shown in drawing 13. . Are preferentially arranged from what it is the search results about the word of a "hamburger" and a "tomato", and is expected to acquire the information that importance is high in the screen shown in the figure. The hyperlink to Web page 20 respectively shown by that position information is embedded at the position information on "the document 1", the "document 3", and the "document 2", and facilities are given to the user of these search results.

[0087]In the analysis of the HTML sauce about Web page 20 in the embodiment described by the above, although it has judged whether the character string is emphasized in Web page 20 based on description of the <FONT> tag and the <B> tag, Based on description of other tags, it may be made to perform this judgment. As an example of a tag employable as the judgment of this emphasis, <I> on which a character string is displayed with an italic character The <U> tag which gives an underline to a tag or a character string, Or at a standard browser, there are a <STRONG> tag etc. which can enable it to make the character string able to pronounce strongly by the voice browser which reads out with a sound the text indicated to the Web page of that on which a character string is only only displayed in a bold letter character. Based on the <FONT> tag in which the FACE attribute for specifying the kind of font used for the display of a character string is specified, It may be made for the character string as which a different font from that by which normal use is carried out to the display sake of the Web page is specified to judge with what is emphasized in the Web page.

[0088]The control program for making it carry out to the computer which has the standard composition which mentioned above the index production processing and retrieval processing which the information site 1 was performing in the embodiment of explained this invention by the above, and the same processing is created, By making the control program read into the computer, and performing it, this invention can be carried out by such computer.

[0089]It is also possible to carry out this invention by computer by making such a control program record on the recording medium which can be read by computer, making the program read from a recording medium to a computer, and performing it. The example of the possible recording medium of reading the control program made to record by computer is shown in drawing 14. The memory storage 502, such as ROM with which the computer 501 is equipped as built-in or external attachment as a recording medium, for example as shown in the figure, and a hard disk drive, or a flexible disk, Portable good signifier recording-media 503 grades, such as MO (magneto-optical disc), CD-ROM, and DVD-ROM, can be used. A recording medium may be the memory storage 506 which is connected with the computer 501 via the network 504 and with which the computer which functions as the program server 505 is provided. In this case, the transmission signal acquired by modulating a subcarrier with the data signal expressing a control program, The control program concerned can be executed now by making it transmit through the network 504 which is a transmission medium from the program server 5055, restoring to the transmission signal received by computer 501, and reproducing a control program.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Brief Description of the Drawings]
[Drawing 1]The information retrieval site which carries out this invention is a figure showing the entire configuration of communication net TEWAKU which provides an information search service.
[Drawing 2]It is a figure showing the detailed composition of an information retrieval site.
[Drawing 3]It is a figure showing an example of a Web page.
[Drawing 4]It is a flow chart which shows the contents of processing of word index production processing.
[Drawing 5]It is a figure showing the example of a HTML filter table.
[Drawing 6]It is a flow chart which shows the contents of processing of the standard setting processing of an index attribute.
[Drawing 7]It is a flow chart which shows the contents of processing of the setting processing of an index attribute.
[Drawing 8]It is a figure explaining the setting processing of an index attribute.
[Drawing 9]It is a figure showing the data structure of an index file.
[Drawing 10]It is a flow chart which shows the contents of processing of retrieval processing.
[Drawing 11]It is a figure showing the example of an index result file.
[Drawing 12]It is a figure showing the situation of sorting of an index result list file.
[Drawing 13]It is a figure showing the example of a screen of the Web page which shows the result of information retrieval.
[Drawing 14]It is a figure showing the example of the possible recording medium of reading the program made to record by computer.
[Description of Notations]
1 Information retrieval site
2a, 2b, and 2c and 2d Offer-of-information site
3a, 3b user terminal
4 Internet

20, 20a, 20b, 20c, 20d Web page

30, 30a, and 30b Browser

100 Research and Data Processing Department

110 Web page collection Management Department

120 Index build Management Department

121 Web page analyzing parts

122 HTML filter table

123 Word extraction Management Department

124 Character emphasis analyzing parts

125 Character string analyzing parts

126 Index registering part

200 Information retrieval Management Department

210 Information retrieval section

211 Search formula storage

220 Search-results Management Department

300 Database manager

310 Index file

320 Search-results file

330 Search-results list file

400 WWW server part

410 HTML preparing part

501 Computer

502 and 506 Memory storage

503 Portable good signifier recording media

504 Network

505 Program server

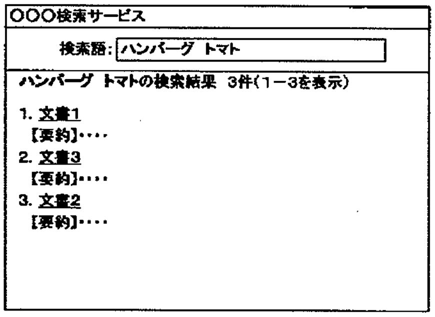
[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Drawing 13]

情報検索の結果を示すWebページの画面例を示す図



○○○検索サービス

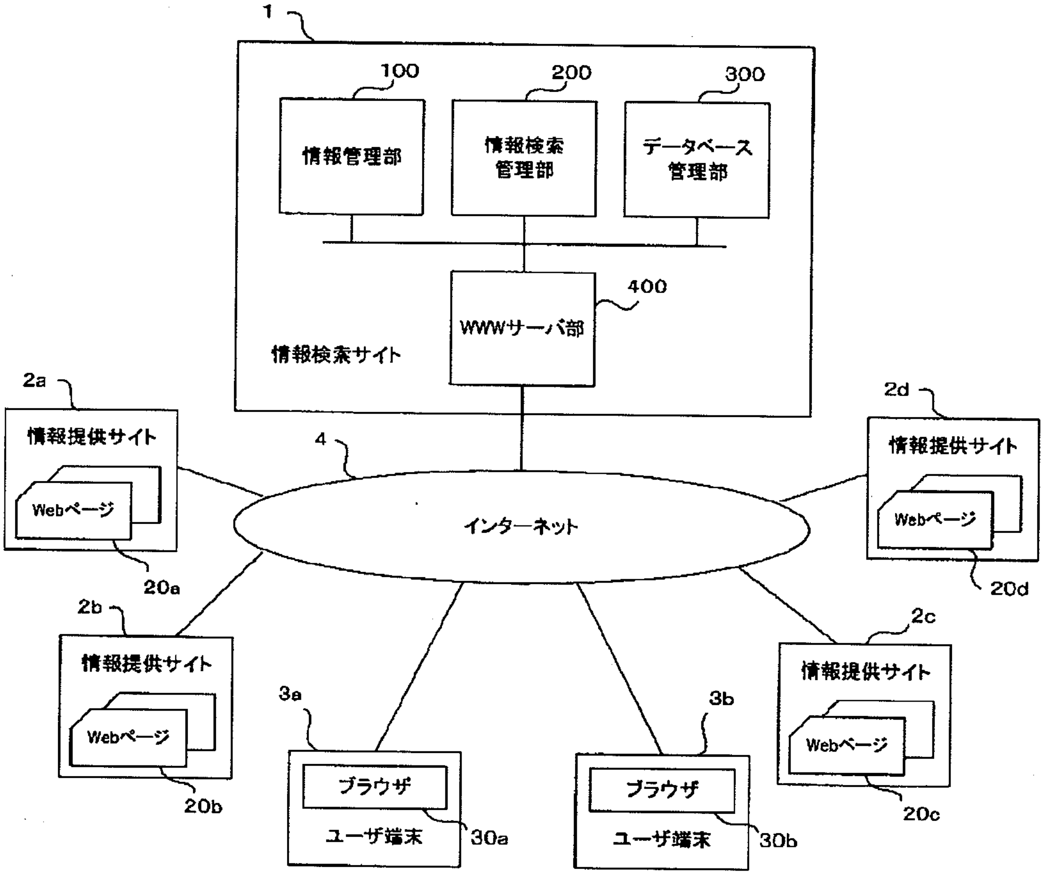
検索語: ハンバーグ トマト

ハンバーグ トマトの検索結果 3件(1－3を表示)

1. 文書1
【要約】・・・・
2. 文書3
【要約】・・・・
3. 文書2
【要約】・・・・

[Drawing 1]

本発明を実施する情報検索サイトが情報検索サービスを
提供する通信ネットワークの全体構成を示す図



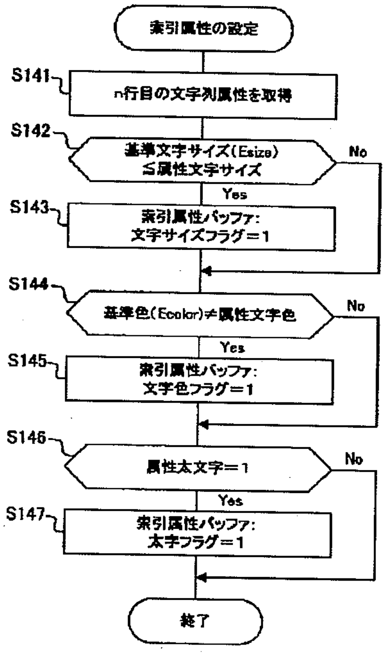
[Drawing 5]

HTMLフィルタテーブルの例を示す図

種別	文字列属性			文字列
	文字ｻｲｽﾞ	色	太字	
STRING	6	赤	1	簡単料理
BR				
STRING	5	青	1	◇ハンバーグのトマトソース煮
BR				
STRING	4	黒		【材料】4人分
BR				
STRING	4	黒		冷凍ハンバーグ:4個
BR				
・・・	・・・	・・・	・・・	・・・
STRING	4	黒		【作り方】
BR				
STRING	4	黒		1. ・・・
BR				
・・・	・・・	・・・	・・・	・・・

[Drawing 7]

索引属性の設定処理の処理内容を
示すフローチャート



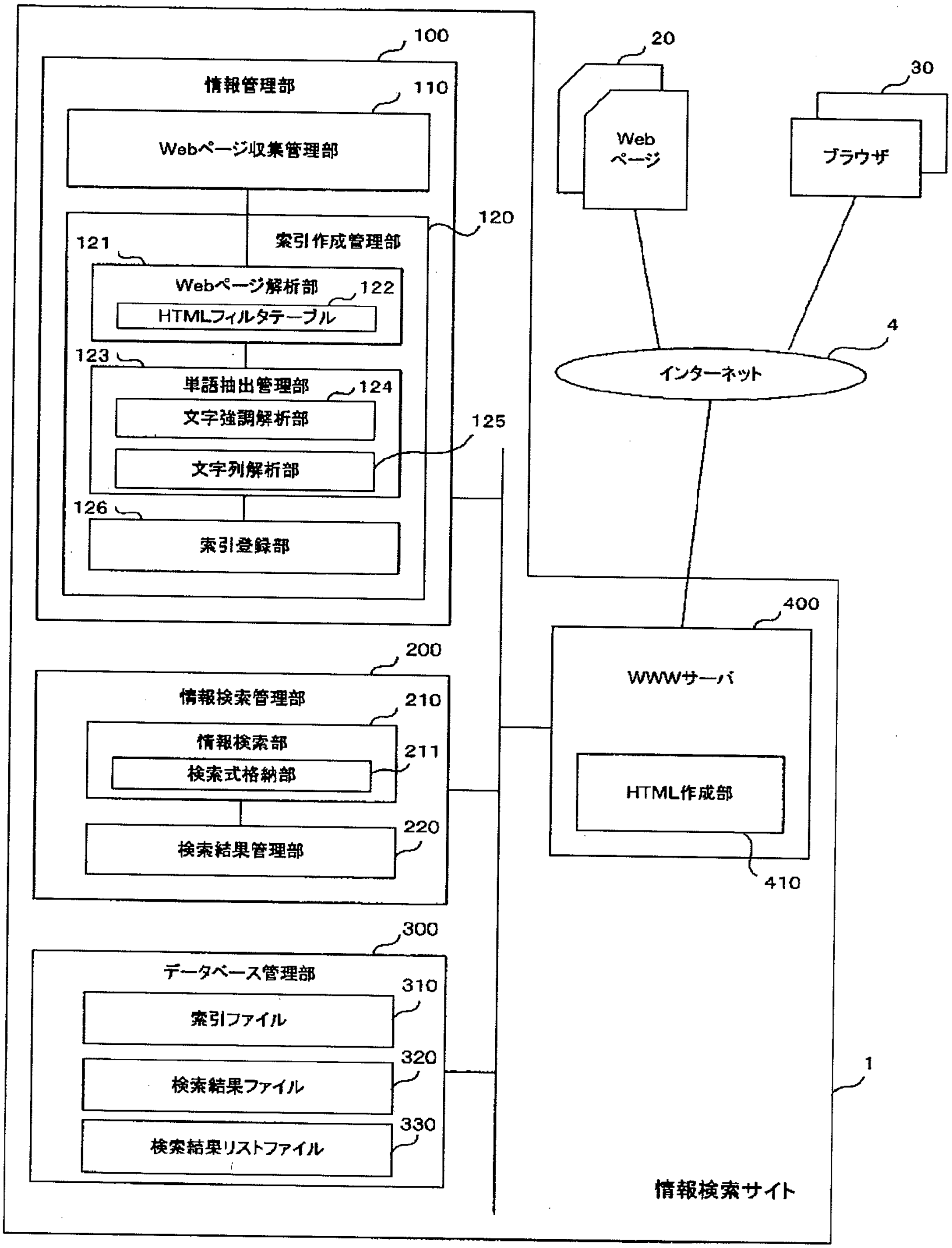
[Drawing 11]

索引結果ファイルの例を示す図

見出し語	位置情報1	要約	属性フラグ1			位置情報2	要約	属性フラグ2			位置情報3	要約	属性フラグ3		
			文字サイズ	色	太字			文字サイズ	色	太字			文字サイズ	色	太字
ハンバーグ	文書1	…	1	1	1	文書2	…	0	0	0	文書3	…	0	1	0
トマト	文書1	…	1	1	1	文書2	…	0	0	0	文書3	…	1	0	1

[Drawing 2]

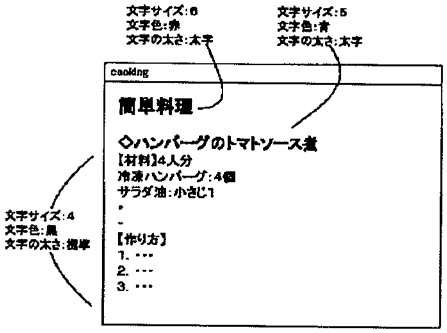
情報検索サイトの詳細構成を示す図



[Drawing 3]

Webページの一例を示す図

(a)ブラウザによって表示されるWebページ画面



cooking

簡単料理

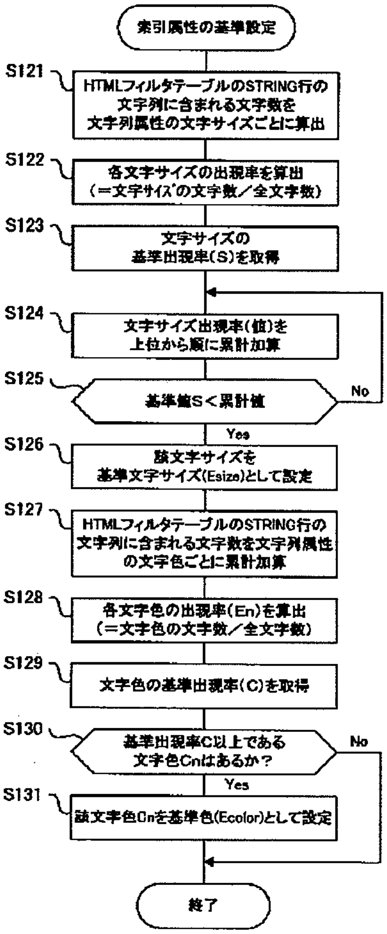
◇ハンバーグのトマトソース煮
【材料】4人分
冷凍ハンバーグ:4個
サラダ油:小さじ1
・
・
【作り方】
1. ・・・
2. ・・・
3. ・・・

(b)HTMLソース

```
<HTML>
<HEAD>
 <TITLE>cooking</TITLE>
</HEAD>
<BODY>
<FONT SIZE="6"COLOR="#FF0000"><B>簡単料理</B></FONT><BR>
<P> </P>
<FONT SIZE="5"COLOR="#0000FF"><B>◇ハンバーグのトマトソース煮
</B></FONT><BR>
【材料】4人分<BR>
冷凍ハンバーグ:4個<BR>
サラダ油:小さじ1<BR>
   ・
   ・
【作り方】<BR>
1. ・・・<BR>
2. ・・・<BR>
3. ・・・<BR>
   ・
   ・
</BODY>
</HTML>
```

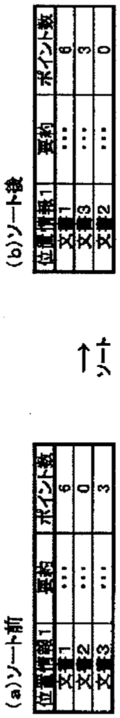
[Drawing 6]

索引属性の基準設定処理の処理内容を示すフローチャート


索引属性の基準設定
S121
HTMLフィルタテーブルのSTRING行の文字列に含まれる文字数を文字列属性の文字サイズごとに算出
S122
各文字サイズの出現率を算出（＝文字サイズの文字数／全文字数）
S123
文字サイズの基準出現率(S)を取得
S124
文字サイズ出現率(値)を上位から順に累計加算
S125
基準値S＜累計値
No
Yes
S126
該文字サイズを基準文字サイズ(Esize)として設定
S127
HTMLフィルタテーブルのSTRING行の文字列に含まれる文字数を文字列属性の文字色ごとに累計加算
S128
各文字色の出現率(En)を算出（＝文字色の文字数／全文字数）
S129
文字色の基準出現率(C)を取得
S130
基準出現率C以上である文字色Cnはあるか？
No
Yes
S131
該文字色Cnを基準色(Ecolor)として設定
終了

[Drawing 12]

検索結果リストファイルのソートの様子を示す図



(a)ソート前

位置情報1	要約	ポイント数
文書1	...	6
文書2	...	0
文書3	...	3

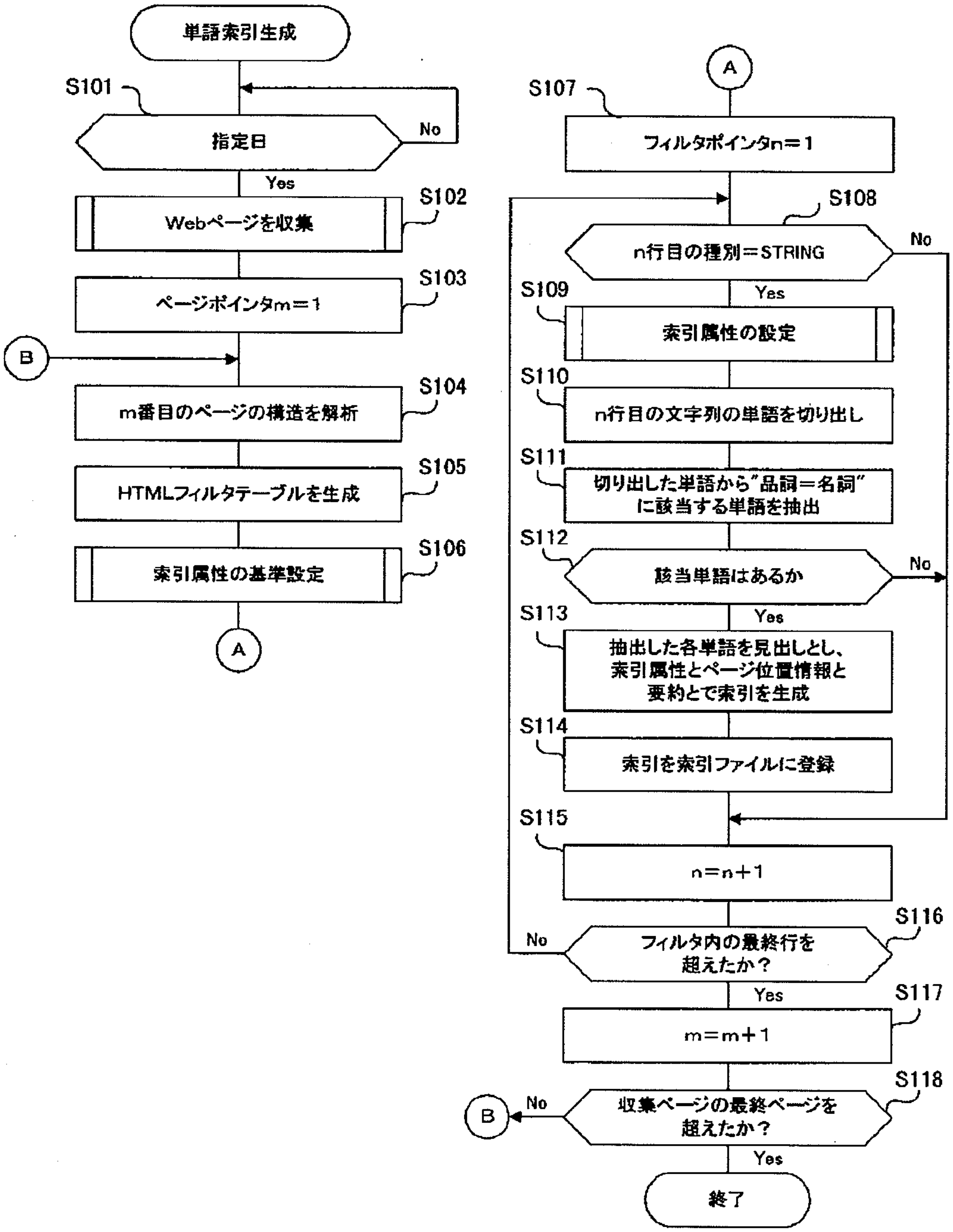
ソート→

(b)ソート後

位置情報1	要約	ポイント数
文書1	...	6
文書3	...	3
文書2	...	0

[Drawing 4]

単語索引生成処理の処理内容を示すフローチャート



[Drawing 8]

索引属性の設定処理を説明する図

(A)文書1

①ページ内文字サイズ出現率テーブル

基準文字サイズ
=最大10%以上

文字サイズ	出現率
7	—
6	3%
5	5%
4	70%
3	12%
2	10%
1	—

②ページ内文字色出現率テーブル

最大出現率文字色
(=90%以上)以外
の文字色

基準色

文字色	出現率
赤	4%
青	6%
黒	90%

(B)文書2

①ページ内文字サイズ出現率テーブル

基準文字サイズ
=最大10%以上

文字サイズ	出現率
7	—
6	—
5	1%
4	—
3	99%
2	—
1	—

②ページ内文字色出現率テーブル

出現率
10%以上

文字色	出現率
赤	50%
青	45%
黒	5%

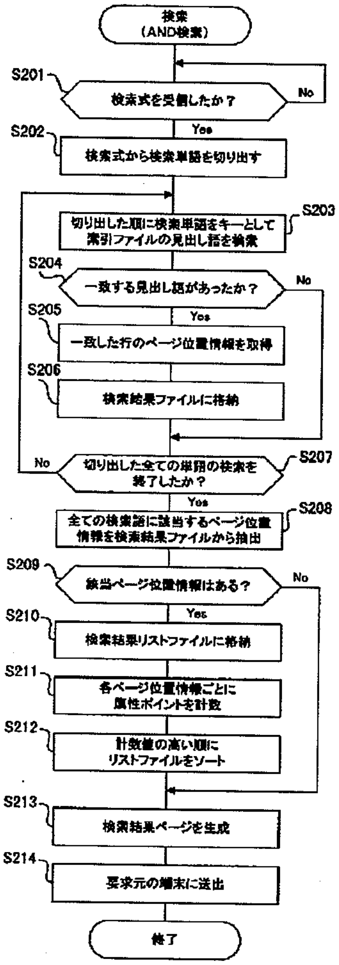
[Drawing 9]

索引ファイルのデータ構造を示す図

見出し語	位置情報1	要約	属性フラグ1			位置情報2	要約	属性フラグ2			…
			文字サイズ	色	太字			文字サイズ	色	太字	
…	…	…	‥	‥	‥	…	…	‥	‥	‥	…
簡単	文書1	…	1	1	1						…
料理	文書1	…	1	1	1	文書2	…	0	0	0	…
ハンバーグ	文書1	…	1	1	1	文書2	…	0	0	0	…
トマト	文書1	…	1	1	1	文書2	…	0	0	0	…
…	…	…	‥	‥	‥	…	…	‥	‥	‥	…
材料	文書1	…	0	0	0						…
人分	文書1	…	0	0	0						…
…	…	…	‥	‥	‥	…	…	‥	‥	‥	…
日記	文書2	…	1	1	1						…
…	…	…	‥	‥	‥	…	…	‥	‥	‥	…

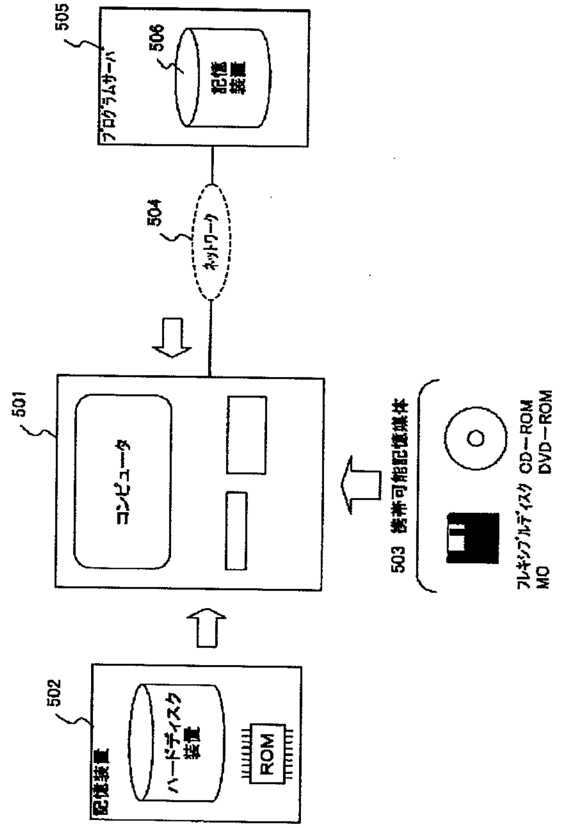
[Drawing 10]

図10の検索処理の処理内容を示すフローチャート


検索
（AND検索）
S201
検索式を受信したか？
No
Yes
S202
検索式から検索単語を切り出す
S203
切り出した順に検索単語をキーとして
索引ファイルの見出し語を検索
S204
一致する見出し語があったか？
No
Yes
S205
一致した行のページ位置情報を取得
S206
検索結果ファイルに格納
No
切り出した全ての単語の検索を
終了したか？
S207
Yes
S208
全ての検索語に該当するページ位置
情報を検索結果ファイルから抽出
S209
該当ページ位置情報はある？
No
Yes
S210
検索結果リストファイルに格納
S211
各ページ位置情報ごとに
属性ポイントを計数
S212
計数値の高い順に
リストファイルをソート
S213
検索結果ページを生成
S214
要求元の端末に送出
終了

[Drawing 14]

記録させた制御プログラムをコンピュータで
読み取ることの可能な記録媒体の例を示す図



[Translation done.]

(19)日本国特許庁（ＪＰ）

(12) **公 開 特 許 公 報**（Ａ）

(11)特許出願公開番号
特開2003－30228
（P2003－30228A）

(43)公開日　平成15年１月31日(2003.1.31)

(51)Int.Cl.^{7}	識別記号	ＦＩ	テーマコード*(参考)
Ｇ０６Ｆ　17/30	２３０	Ｇ０６Ｆ　17/30	２３０Ｚ　５Ｂ０７５
	３４０		３４０Ｚ
	３５０		３５０Ｚ
	４１９		４１９Ｂ

審査請求　未請求　請求項の数７　ＯＬ　（全 17 頁）

(21)出願番号　特願2001－212555(P2001－212555)

(22)出願日　平成13年７月12日(2001.7.12)

(71)出願人　000001443
カシオ計算機株式会社
東京都渋谷区本町１丁目６番２号

(72)発明者　寺田　俊仁
東京都羽村市栄町３丁目２番１号　カシオ計算機株式会社羽村技術センター内

(74)代理人　100093632
弁理士　阪本　紀康

Ｆターム(参考)　5B075 ND36 NR03 PQ75 PR08 PR10
QM10

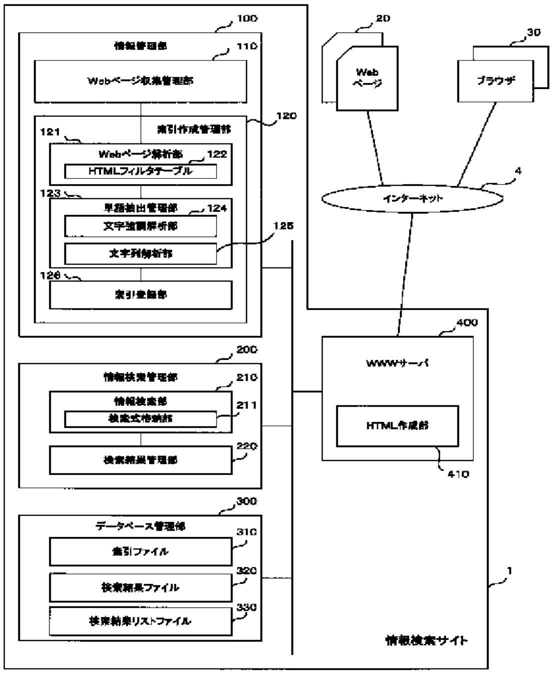
(54)【発明の名称】　情報検索システム、情報検索方法、及びプログラム

(57)【要約】

【課題】　検索エンジンが提供する情報検索の結果を情報検索者の検索目的に対してより適切なものにする。

【解決手段】　文字強調解析部１２４はＷｅｂページ２０に含まれている文字列に与えられている強調を示す属性を取得する。索引登録部１２６は、該文字列から文字列解析部１２５によって抽出された単語に、Ｗｅｂページ２０の論理的な位置情報と該単語の抽出元の文字列についての強調属性とを対応付けて索引ファイル３１０に登録する。情報検索部２１０は検索対象を表す単語に対応付けられている位置情報を取得する。検索結果管理部２２０は、取得された位置情報をソートして、該位置情報の検索の対象とした単語に強調属性の対応付けられているものが優先されるようにする。ソートされた位置情報を表すＨＴＭＬファイルがＨＴＭＬ作成部４１０で作成されてインターネット４に送出される。

情報検索サイトの詳細構成を示す図



【特許請求の範囲】

【請求項１】　通信ネットワーク上で公開されている文書情報に含まれている単語と、該通信ネットワーク上の論理的な位置を示す情報であって該単語を含む情報が存在する文書情報位置を示す位置情報とを対応付けてなる索引ファイルを用意し、検索の対象を表す単語に基づいて該索引ファイルを検索して該検索対象を表す単語に対応している位置情報を提示するシステムであって、

前記文書情報に含まれている文字列に与えられている属性であって強調を示す強調属性を取得する強調属性取得手段と、

前記文字列から単語を抽出する抽出手段と、

前記抽出手段によって抽出された単語に、該単語についての前記位置情報と該単語の抽出元の文字列に与えられている前記強調属性とを対応付けて前記索引ファイルに登録する登録手段と、

前記検索対象を表す単語に基づいて前記索引ファイルの検索を行なって該単語に対応付けられている位置情報を該索引ファイルから取得する検索手段と、

前記検索手段によって取得された位置情報のうち、前記索引ファイルにおいて該検索手段が該位置情報の検索の対象とした単語に前記強調属性が対応付けられているものを優先して該位置情報を提示する提示手段と、

を有することを特徴とする情報検索システム。

【請求項２】　前記文書情報に含まれている文字列に与えられている属性は、該文字列を表示するときに用いられる文字の大きさを示す属性、若しくは該文字列を表示するために用いられる文字の色彩を示す属性であることを特徴とする請求項１に記載の情報検索システム。

【請求項３】　前記強調属性取得手段は、

前記文書情報に含まれている文字列に与えられている属性についての該文書情報における出現頻度を該属性毎に算出する出現頻度算出手段と、

前記属性が前記強調属性であるか否かを判別する基準を該属性毎の出現頻度に基づいて設定する強調属性設定手段と、

前記基準に基づいて、前記文書情報に含まれている文字列に与えられている属性が前記強調属性であるか否かを判別する強調属性判別手段と、

を有することを特徴とする請求項１又は２に記載の情報検索システム。

【請求項４】　前記強調属性取得手段は、前記文書情報に含まれている文字列に与えられている属性が該文字列を表示するために用いられる文字を太字とする旨を示しているときには、該属性を前記強調属性であるとみなすことを特徴とする請求項１に記載の情報検索システム。

【請求項５】　前記提示手段は、前記検索手段によって取得された位置情報のうち、前記検索ファイルにおいて該検索手段が該位置情報の検索の対象とした単語に対応付けられている前記強調属性の数が多いものほど優先して該位置情報を提示することを特徴とする請求項１に記載の情報検索システム。

【請求項６】　通信ネットワーク上で公開されている文書情報に含まれている単語と、該通信ネットワーク上の論理的な位置を示す情報であって該単語を含む情報が存在する文書情報位置を示す位置情報とを対応付けてなる索引ファイルを用意し、検索の対象を表す単語に基づいて該索引ファイルを検索して該検索対象を表す単語に対応している位置情報を提示する方法であって、

前記文書情報に含まれている文字列に与えられている属性であって強調を示す強調属性を取得し、

前記文字列から単語を抽出し、

前記文字列から抽出された単語に、該単語についての前記位置情報と該単語の抽出元の文字列に与えられている前記強調属性とを対応付けて前記索引ファイルに登録し、

前記検索対象を表す単語に基づいて前記索引ファイルの検索を行なって該単語に対応付けられている位置情報を該索引ファイルから取得し、

前記検索によって取得された位置情報のうち、前記検索ファイルにおいて該位置情報の検索の対象とした単語に前記強調属性が対応付けられているものを優先して該位置情報を提示する、

ことを特徴とする情報検索方法。

【請求項７】　コンピュータに実行させることにより、通信ネットワーク上で公開されている文書情報に含まれている単語と該通信ネットワーク上の論理的な位置を示す情報であって該単語を含む情報が存在する文書情報位置を示す位置情報とを対応付けてなる索引ファイルを用意する処理と、検索の対象を表す単語に基づいて該索引ファイルを検索して該検索対象を表す単語に対応している位置情報を提示する処理とを該コンピュータに行なわせるためのプログラムであって、

前記文書情報に含まれている文字列に与えられている属性であって強調を示す強調属性を取得する処理と、

前記文字列から単語を抽出する処理と、

前記文字列から抽出された単語に、該単語についての前記位置情報と該単語の抽出元の文字列に与えられている前記強調属性とを対応付けて前記索引ファイルに登録する処理と、

前記検索対象を表す単語に基づいて前記索引ファイルの検索を行なって該単語に対応付けられている位置情報を該索引ファイルから取得する処理と、

前記検索によって取得された位置情報のうち、前記検索ファイルにおいて該位置情報の検索の対象とした単語に前記強調属性が対応付けられているものを優先して該位置情報を提示する処理と、

をコンピュータに行なわせるためのプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、情報を検索する技術に関し、特に、検索の要求に対し、その要求により適切に合致した情報を提供できるようにする技術に関する。

【０００２】

【従来の技術】近年、インターネットの普及により、インターネット上のＷＷＷ（World WideWeb）システムで提供されているＷｅｂページの数は爆発的に増え続けている。また、インターネット上では、この膨大な情報の中から目的とする情報を検索するサービスを提供する検索エンジンが多数開設されている。

【０００３】検索エンジンがネット上の情報を収集する方式のひとつとしてロボット型と称されているものがある。ロボット型の検索エンジンでは、スパイダあるいはクローラなどと呼ばれるロボットプログラムが定期的に起動されて、インターネット上で公開されているＷｅｂページを表現しているＨＴＭＬ（HyperText Markup Language ）ファイルの自動収集が行なわれる。情報検索が行なわれるときには、検索エンジンを利用する情報検索者が目的とする情報に関係の深いキーワードを検索サイトに与えることにより、収集されたファイルからそのキーワードが含まれたものを抽出する処理が行なわれ、そのキーワードの含まれているＷｅｂページのリストが、そのＷｅｂページについてのインターネット上における論理的な位置を示す情報と共に、検索結果として情報検索者に提供される。

【０００４】

【発明が解決しようとする課題】一般に、ロボット型の検索エンジンは情報の収集から検索結果の提供に至るまでの全ての処理をコンピュータで自動的に行なっており、そこには人間の判断による情報の操作が介在しないので、収集された情報の属するジャンルやその情報の質についての整理がなされていない。そのため、情報の検索の際に単なるキーワードの一致による検索を行なっていたのでは、重要な情報を含むＷｅｂページが検索結果に埋もれてしまったり、あるいは、いわゆる検索ノイズ、すなわち有用性の低い情報しか含まれていないＷｅｂページばかり検索結果に多く含まれてしまったりする場合が少なくなかった。

【０００５】以上の問題を鑑み、検索エンジンが提供する情報検索の結果を情報検索者の検索目的に対してより適切なものにすることが本発明が解決しようとする課題である。

【０００６】

【課題を解決するための手段】本発明は、通信ネットワーク上で公開されている文書情報に含まれている単語と、該通信ネットワーク上の論理的な位置を示す情報であって該単語を含む情報が存在する文書情報位置を示す位置情報とを対応付けてなる索引ファイルを用意し、検索の対象を表す単語に基づいて該索引ファイルを検索して該検索対象を表す単語に対応している位置情報を提示するシステムまたは方法を前提とする。

【０００７】そして、本発明の態様のひとつである情報検索システムは、前記文書情報に含まれている文字列に与えられている属性であって強調を示す強調属性を取得する強調属性取得手段と、前記文字列から単語を抽出する抽出手段と、前記抽出手段によって抽出された単語に、該単語についての前記位置情報と該単語の抽出元の文字列に与えられている前記強調属性とを対応付けて前記索引ファイルに登録する登録手段と、前記検索対象を表す単語に基づいて前記索引ファイルの検索を行なって該単語に対応付けられている位置情報を該索引ファイルから取得する検索手段と、前記検索手段によって取得された位置情報のうち、前記索引ファイルにおいて該検索手段が該位置情報の検索の対象とした単語に前記強調属性が対応付けられているものを優先して該位置情報を提示する提示手段と、を有するように構成することによって前述した課題を解決する。

【０００８】ここで、前記文書情報に含まれている文字列に与えられている属性は、例えば該文字列を表示するときに用いられる文字の大きさを示す属性、あるいは該文字列を表示するために用いられる文字の色彩を示す属性である。強調が指示されている単語を含む文書情報はその単語に関し重要度の高い情報が含まれている可能性が高いと考えることができる。従って、上述した構成によれば、検索条件を示す単語に合致する文書情報が通信ネットワークに複数公開されているときに、その単語が強調されているため重要度が高いと考えられる文書情報についての位置情報が優先されるようにして提示されるので、情報検索の結果が情報検索者の検索目的に対してより適切なものとなる。

【０００９】なお、上述した本発明に係る情報検索システムにおいて、前記強調属性取得手段は、前記文書情報に含まれている文字列に与えられている属性についての該文書情報における出現頻度を該属性毎に算出する出現頻度算出手段と、前記属性が前記強調属性であるか否かを判別する基準を該属性毎の出現頻度に基づいて設定する強調属性設定手段と、前記基準に基づいて、前記文書情報に含まれている文字列に与えられている属性が前記強調属性であるか否かを判別する強調属性判別手段と、を有するように構成してもよい。

【００１０】この構成によれば、文書情報において特異な属性が与えられている文字列はその文書情報において強調されていると判断することができるようになるので、この文字列に含まれる単語にとってこの文書情報は重要度が高いものとして索引ファイルに登録できるようになる。

【００１１】また、前述した本発明に係る情報検索システムにおいて、前記強調属性取得手段は、前記文書情報に含まれている文字列に与えられている属性が該文字列

を表示するために用いられる文字を太字とする旨を示しているときには、該属性を前記強調属性であるとみなすようにしてもよい。

【００１２】この構成によれば、文書情報において文字列の強調を示すために広く行なわれている太字による表示を行なう旨の属性については直ちに強調属性と判定することができるようになる。また、前述した本発明に係る情報検索システムにおいて、前記提示手段は、前記検索手段によって取得された位置情報のうち、前記検索ファイルにおいて該検索手段が該位置情報の検索の対象とした単語に対応付けられている前記強調属性の数が多いものほど優先して該位置情報を提示するようにしてもよい。

【００１３】この構成によれば、ある文字列に与えられている強調属性の数が多いものほど、この文字列に含まれる単語にとってこの文書情報は重要度がより高いものとして索引ファイルに登録できるようになり、情報検索の結果が情報検索者の検索目的に対して更に適切なものとなる。

【００１４】本発明の別の態様のひとつである情報検索方法は、前記文書情報に含まれている文字列に与えられている属性であって強調を示す強調属性を取得し、前記文字列から単語を抽出し、前記文字列から抽出された単語に、該単語についての前記位置情報と該単語の抽出元の文字列に与えられている前記強調属性とを対応付けて前記索引ファイルに登録し、前記検索対象を表す単語に基づいて前記索引ファイルの検索を行なって該単語に対応付けられている位置情報を該索引ファイルから取得し、前記検索によって取得された位置情報のうち、前記検索ファイルにおいて該位置情報の検索の対象とした単語に前記強調属性が対応付けられているものを優先して該位置情報を提示することにより、前述した本発明に係る情報検索システムと同様の作用・効果が得られる。

【００１５】なお、上述した本発明に係る情報検索方法と同様の手順からなる処理をコンピュータに行なわせるためのプログラムでも、そのプログラムをコンピュータに実行させることによって前述した課題を解決することができる。

【００１６】

【発明の実施の形態】以下、本発明の実施の形態を図面に基づいて説明する。図１は本発明を実施する情報検索サイトが情報検索サービスを提供する通信ネットワークの全体構成を示す図である。

【００１７】図１において、情報検索サイト１、情報提供サイト２ａ、２ｂ、２ｃ、２ｄ、及びユーザ端末３ａ、３ｂはいずれも通信ネットワークであるインターネット４に接続されており、相互にデータの授受を行なうことができる。情報検索サイト１はユーザ端末３ａ及び３ｂにロボット検索型の情報検索サービスを提供するWWWサーバシステムであり、情報管理部１００、情報検索管理部２００、情報データベース管理部３００、及びWWWサーバ管理部４００を備えて構成されている。

【００１８】情報管理部１００はインターネット４上に公開されている情報の自動収集を行ない、収集された情報を情報データベース管理部３００に蓄積する。情報検索管理部２００は、インターネット４を介して送られてくる情報検索の要求に応じ、情報データベース管理部３００に蓄積されている情報の検索を行ない、その検索の結果を要求元に返送する。

【００１９】情報データベース管理部３００では情報管理部１００によって収集された情報の蓄積、及び情報検索管理部２００による情報の検索が行なわれる。WWWサーバ部４００は、インターネット４を介して送られてくる収集された情報を情報管理部１００に転送する処理、インターネット４を介して送られてくる情報検索の要求を情報検索管理部２００に転送する処理、及び情報検索管理部２００から送られてくる情報検索の結果を示す情報が表されているＷｅｂページの送出の処理が行なわれる。

【００２０】情報提供サイト２ａ、２ｂ、２ｃ、及び２ｄは、それぞれＷｅｂページ２０ａ、２０ｂ、２０ｃ、及び２０ｄをインターネット４上で公開するWWWサーバシステムである。なお、図１においては４つの情報提供サイトを示しているが、インターネット４に接続される情報提供サイトの数は任意でよい。

【００２１】ユーザ端末３ａ及び３ｂは、それぞれ情報提供サイト２ａ、２ｂ、２ｃ、及び２ｄや情報検索サイト１から提供されるＷｅｂページを閲覧するソフトウェアであるブラウザ３０ａ及び３０ｂを実行可能なコンピュータであり、インターネット４上で公開されている情報の検索を情報検索サイト１へ依頼する情報検索者によって操作される。なお、図１においては２つのユーザを示しているが、インターネット４に接続されるユーザ端末の数も任意でよい。

【００２２】なお、これらの情報検索サイト１、情報提供サイト２ａ、２ｂ、２ｃ、及び２ｄ、ユーザ端末３ａ及び３ｂは、いずれも標準的な構成を有するコンピュータ、すなわち、制御プログラムを実行することで各構成要素を制御するＣＰＵと、ＲＯＭやＲＡＭ及び磁気記憶装置などからなり、ＣＰＵに各構成要素を制御させる制御プログラムの記憶やＣＰＵが制御プログラムを実行する際のワークエリアあるいは各種データの記憶領域として使用される記憶部と、ユーザによる操作に対応する各種のデータが取得される入力部と、ディスプレイなどに各種のデータを提示してユーザに通知する出力部と、ネットワークに接続するためのインタフェース機能を提供するＩ／Ｆ部とを備えるコンピュータを用いて構成することもできる。

【００２３】次に図２について説明する。同図は本発明を実施する図１における情報検索サイト１の詳細構成を

示している。図２に示すように、情報管理部１００はＷｅｂページ収集管理部１１０及び索引生成部１２０を備えて構成され、情報管理検索部２００は情報検索部２１０及び検索結果管理部２２０を備えて構成され、データベース管理部３００は索引ファイル３１０、検索結果ファイル３２０、及び検索結果リストファイル３３０を備えて構成され、そしてＷＷＷサーバ４００はＨＴＭＬ作成部４１０を備えて構成される。

【００２４】Ｗｅｂページ収集管理部１１０はインターネット４上で公開されているＷｅｂページ２０の自動収集を定期的に行なう。索引作成管理部１２０は、Ｗｅｂページ収集管理部１１０によって収集されたＷｅｂページ２０の位置情報、すなわちＷｅｂページ２０が存在するインターネット４上の論理的な位置を示す位置情報を引くことのできる索引となるレコードを作成して索引ファイル３１０に登録する。索引作成管理部１２０はＷｅｂページ解析部１２１、単語抽出管理部１２３、索引登録部１２６を備えて構成されている。

【００２５】Ｗｅｂページ解析部１２１はＷｅｂページ２０の解析を行なってＷｅｂページ２０を表現しているＨＴＭＬファイルの本文に記述されている各ＨＴＭＬタグをレコードの単位とするＨＴＭＬフィルタテーブル１２２を作成する。単語抽出管理部１２３では、ＨＴＭＬフィルタテーブル１２２に示されている文字列においてそれらがＷｅｂページ２０の画面として表示されたときに強調されているとみなし得る文字書式の条件を解析する処理が文字強調解析部１２４で行なわれ、また、ＨＴＭＬフィルタテーブル１２２に示されている文字列の解析が文字列解析部１２５で行なわれてその文字列から単語が抽出される。

【００２６】索引登録部１２６は、文字列解析部１２５によって抽出された単語を見出し語とし、その見出し語に、Ｗｅｂページ２０についてのインターネット４上における論理的な位置情報と、その単語が含まれていたＷｅｂページ２０の要約と、Ｗｅｂページ２０でその単語に設定されていた書式が文字強調解析部１２４での解析によって得られた文字書式の条件に合致することを示す属性フラグとを対応付けた索引レコードを索引ファイル３１０に登録する。

【００２７】情報検索部２１０は、インターネット４に接続されているいずれかのユーザ端末で実行されているブラウザ３０の制御によってそのユーザ端末から送られてくる情報検索の要求をＷＷＷサーバ部４００から取得し、その情報検索の条件を示す検索式をその要求から取り出して検索式格納部２１１に格納する。そして、索引ファイル３１０を検索してその検索式に示されている単語（キーワード）が見出しとなっている索引レコードを取得して検索結果ファイル３２０に格納する。

【００２８】検索結果管理部２２０は、情報検索部２１０による検索が完了すると、検索結果ファイル３２０に格納されている索引データに示されている位置情報及び要約と、その索引レコードでその位置情報に対応して付されている属性フラグの合計数とを検索結果リストファイル３３０に格納する。そして、検索結果リストファイル３３０に格納された位置情報を合計数に従ってソートする。

【００２９】ＨＴＭＬ作成部４１０は、検索結果リストファイル３３０に格納されているソートされた位置情報からなる検索結果リストを受け取ってその検索結果リストが表現されるＷｅｂページを表現するＨＴＭＬファイルを作成する。作成されたＨＴＭＬファイルはブラウザ３０が実行されているユーザ端末へ宛ててＷＷＷサーバ部４００によりインターネット４に送出される。

【００３０】次に、情報検索サイト１の有する情報管理部１００において行なわれる、Ｗｅｂページの収集及び索引の生成の処理の詳細について説明する。図３は、インターネット４に公開されていて情報検索サイト１によって収集されるＷｅｂページ２０の例を示している。同図において、（ｂ）に示すＨＴＭＬソースがブラウザ３０によって閲覧されると同図（ａ）に示す画面が表示される。

【００３１】ここで図４について説明する。同図は情報管理部１００で実行される索引生成処理の処理内容を示すフローチャートである。この処理が実行されることによって、Ｗｅｂページの収集及び索引の生成が情報管理部１００で行なわれる。まず、Ｓ１０１において、現在の日付が、予め指定されているＷｅｂページ２０の収集指定日であるか否かがＷｅｂページ収集管理部１１０で判別され、この判定結果がＹｅｓ、すなわち現在がその指定日となったときにのみ、処理がＳ１０２に進む。この日付の指定の仕方は任意であるが、例えば毎月の月末最終日などといった指定が行なわれる。

【００３２】Ｓ１０２ではインターネット４上で公開されているＷｅｂページ２０の巡回・収集の処理がＷｅｂページ収集管理部１１０によって行なわれる。この巡回・収集の手法は周知のロボット型検索エンジンで従来から行なわれているものをそのまま利用すればよい。

【００３３】Ｓ１０３では、Ｗｅｂページ解析部１２１において、前ステップで収集された多数のＷｅｂページ２０を１ページずつ指し示すためのポインタであるページポインタとして使用される変数ｍに初期値１が代入される。Ｓ１０４ではＳ１０２の処理によって収集されたＷｅｂページ２０におけるページポインタｍの現在の値で指し示されるページの構造がＷｅｂページ解析部１２１によって解析され、続くＳ１０５においてＨＴＭＬフィルタテーブル１２２がＷｅｂページ解析部１２１によって生成される。

【００３４】図３に示したＷｅｂページから生成されるＨＴＭＬフィルタテーブルを図５に示す。Ｗｅｂページ解析部１２１によって図３（ｂ）に示したＨＴＭＬソー

スが解析されると図５に示すＨＴＭＬフィルタテーブルが生成される。Ｓ１０４の処理内容について図３（ｂ）を参照しながら更に説明すると、Ｗｅｂページ解析部１２１では、解析対象のＨＴＭＬソースの本文、すなわち<;BODY>;の開始タグと終了タグとの間に挟まれている部分が解析の対象とされ、その部分における<;BR>;タグ（改行タグ）が文の区切りとされてその本文中に含まれる各文の構造が解析される。

【００３５】図３（ｂ）に示すＨＴＭＬソースから図５に示すＨＴＭＬフィルタテーブルが作成される様子について説明すると、まず、ＨＴＭＬソースにおける本文の記述部分である<;BODY>;タグの開始タグと終了タグとの間に挟まれている部分、すなわち<;BODY>;タグと<;/BODY>; タグとに挟まれている部分がＳ１０４の処理による解析の対象とされる。

【００３６】ここで、まず、解析対象の部分における最初の<;BR>;タグが記述されている箇所までの部分、すなわち、「 <;FONT SIZE="6"COLOR="#FF0000">;<;B>;簡単料理<;/B>;<;/FONT>; 」なる部分が解析される。ここで、<;FONT SIZE="6"COLOR="#FF0000">;なるタグは、<;FONT>;の開始タグと終了タグとの間に挟まれている部分に記述されている「簡単料理」なる文字列について、文字サイズを「６」なる大きさとし且つ文字色を「＃ＦＦ００００」なる数値で示される色として表示することを示している。なお、この数値で示される色は赤色である。

【００３７】また、<;B>; なるタグは、<;FONT>;の開始タグと終了タグとの間に挟まれている部分に記述されている「簡単料理」なる文字列を太字で表示することを示している。Ｓ１０４による解析処理は上述した解析を行なうものであり、この後に実行されるＳ１０５の処理によって、ＨＴＭＬページフィルタ１２２にはこの解析結果の内容を意味する図５の第１行目に示すレコードが格納される。このレコードを説明すると、「文字列」の欄には「簡単料理」が格納され、「種別」の欄には「簡単料理」が文字列であることを示す「ＳＴＲＩＮＧ」が格納され、「文字列属性」の欄には「簡単料理」の文字列について「文字サイズ」を「６」、「色」を「赤」として表示することを示す属性が格納される。さらに「文字列属性」における「太字」の欄にフラグ「１」が格納されることによって「簡単料理」の文字列を太字で表示することが示されていたことが分かる。そして、その次に「種別」が「ＢＲ」であることのみを内容とする、<;BR>;タグの存在を示すレコードがＨＴＭＬページフィルタ１２２の第２行目に格納される。

【００３８】以下、図３（ｂ）に示すＨＴＭＬソースの本文部分における残りの部分の解析も同様に行なわれ、こうして図５に示すＨＴＭＬページフィルタ１２２が生成される。なお、<;FONT>;タグにおけるSIZE属性は、"+1"や"-1"などというように、通常の表示文字サイズに対する相対値で示されている場合もあるが、そのようなときにはその相対値を文字サイズとして登録するようにし、文字サイズの指定がなされていない文字列については"0" を文字サイズとして登録する。

【００３９】図４の説明へ戻り、Ｓ１０６ではＨＴＭＬページフィルタ１２２に格納された文字列において、索引属性の基準の設定、すなわち強調表示を行なうことが設定されているとみなせる文字列属性の基準を解析してその基準を設定する処理が文字強調解析部１２４で行なわれる。この処理の詳細は後述する。

【００４０】Ｓ１０７では文字列解析部１２５によってＨＴＭＬフィルタテーブル１２２の各行（レコード）を順番にひとつずつ指定するためのポインタであるフィルタポインタとして使用される変数ｎに初期値１が代入される。Ｓ１０８では、上述したフィルタポインタｎによって指定される行に示されている文字列の種別を示すデータが「ＳＴＲＩＮＧ」であるか否かが文字列解析部１２５によって判別され、この判別の結果がＹｅｓならばＳ１０９に処理が進み、ＮｏならばＳ１１５に処理が進む。

【００４１】Ｓ１０９では、フィルタポインタｎによって指定される行に示されている文字列に、Ｓ１０６の処理によって設定された基準に基づいて索引属性を設定する処理が文字列解析部１２５によって行なわれる。この処理の詳細も後述する。その後、Ｓ１１０においてフィルタポインタｎによって指定される行に示されている文字列から単語を切り出す処理が文字列解析部１２５によって行なわれ、Ｓ１１１において、その切り出された単語からその品詞が名詞である単語を抽出する処理が文字列解析部１２５によって続けて行なわれる。

【００４２】なお、Ｓ１１０における単語の切り出しの処理には周知の方式を採用する。この周知の方式としては、例えばいわゆる形態素解析を利用し、切り出した単語の品詞と活用形を各種の辞書を用いてその単語の標準形を取得してその標準形の単語を文字列から切り出した単語とする方式や、文字列の切り出しをその文字列の先頭から１文字ずつずらしながら順に長さＮの語を機械的に切り出すいわゆるＮグラム方式などがある。

【００４３】Ｓ１１２では、上述したＳ１１１の処理によって抽出された単語が存在したか否かが文字列解析部１２５によって判別され、この判定結果がＹｅｓならばＳ１１３に処理が進み、ＮｏならばＳ１１５に処理が進む。Ｓ１１３では、前述したＳ１１１の処理によって抽出された単語を見出しとし、その単語が含まれていたページの位置情報と、そのページに記載されている文章の要約と、その単語が含まれていた文字列に対してＳ１０９の処理によって設定された文字属性とをその見出しの単語に対応付けた索引が索引登録部１２６で生成され、続くＳ１１４の処理によって生成された索引が索引ファイル３１０に登録される。

【００４４】Ｓ１１５では、文字列解析部１２５によっ

て前述したフィルタポインタｎの指示が１だけ進められる。Ｓ１１６では、フィルタポインタｎの現在の数値によって指定される行がＨＴＭＬフィルタテーブル１２２に存在する最終の行を超えてしまったか否かが文字列解析部１２５によって判別され、この判別結果がＹｅｓならば処理がＳ１１７に進み、ＮｏならばＳ１０８へ処理が戻って上述した処理が繰り返される。

【００４５】Ｓ１１７では、Ｗｅｂページ解析部１２１によって前述したページポインタｍの指示が１だけ進められる。Ｓ１１８では、ページポインタｍの現在の数値によって指定されるページがＷｅｂページ収集管理部１１０によって収集されたＷｅｂページ２０の最終のページを超えてしまったか否かがＷｅｂページ解析部１２１によって判別され、この判別の結果がＹｅｓならばこの索引生成処理が終了する。一方、この判別処理の結果がＮｏならばＳ１０４へ処理が戻って上述した処理が繰り返される。

【００４６】以上までの処理が索引生成処理である。次に、上述した索引生成処理におけるＳ１０６の処理として文字強調解析部１２４で行なわれる索引属性の基準設定処理の詳細について説明する。図６は索引属性の基準設定処理の処理内容を示すフローチャートである。

【００４７】まず、Ｓ１２１において、ＨＴＭＬフィルタテーブル１２２における「種別」が「ＳＴＲＩＮＧ」である行の文字列の文字数の合計が、「文字列属性」における文字サイズ毎に算出される。次に、Ｓ１２２において、各文字サイズの出現率、すなわちＨＴＭＬフィルタテーブル１２２に示されている全ての文字列の文字数に対する各文字サイズの文字数の割合が算出される。

【００４８】Ｓ１２３では、予め設定されている文字サイズについての基準出現率Ｓが取得される。Ｓ１２４では、Ｓ１２２の処理によって算出された出現率が文字サイズの大きい順に累計加算され、続くＳ１２５において、その累計値が基準値Ｓを上回ったか否かが判別される。そして、この判別結果がＹｅｓとなったときに処理がＳ１２６に進む。一方、この判別結果がＮｏである間は、Ｓ１２４の処理が繰り返される。

【００４９】Ｓ１２６では、前ステップの判別処理の結果がＹｅｓとなったときにその直前のＳ１２４の処理で加算された出現率に対応する文字サイズよりもひとつ大きい文字サイズが基準文字サイズＥｓｉｚｅとして設定される。後述する索引属性の基準設定処理においては、この基準文字サイズＥｓｉｚｅ以上の文字サイズの属性の設定されている文字列が、Ｗｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５０】Ｓ１２７では、ＨＴＭＬフィルタテーブル１２２における「種別」が「ＳＴＲＩＮＧ」である行に示されている全ての文字列の文字数の合計が、「文字列属性」における文字色毎に算出される。Ｓ１２８では、各文字色の出現率、すなわちＨＴＭＬフィルタテーブル１２２に示されている全ての文字列の文字数に対する各文字色の文字数の割合が算出される。

【００５１】Ｓ１２９では、予め設定されている文字色についての基準出現率Ｃが取得される。Ｓ１３０では、出現率が基準出現率Ｃ以上である文字色Ｃｎが存在するか否かが判別され、この判別結果がＹｅｓのときにのみ、Ｓ１３１において、この文字色Ｃｎが基準色Ｅｃｏｌｏｒとして設定される。後述する索引属性の基準設定処理においては、この基準色Ｅｃｏｌｏｒとは異なる文字色の属性の設定されている文字列が、Ｗｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５２】Ｓ１３０及びＳ１３１の処理を終えるとこの索引属性の基準設定処理が終了し、前述した図４へと処理が戻る。以上までの処理が索引属性の基準設定処理である。次に、上述した索引属性の基準設定処理を図８の例を用いて更に説明する。

【００５３】図８（Ａ）は、文書１なるＷｅｂページに含まれる文字の出現率を示しており、①はＳ１２２までの処理によって得られた各文字サイズの出現率を示すテーブル、②はＳ１２８までの処理によって得られた各文字色の出現率を示すテーブルである。

【００５４】今、Ｓ１２３の処理によって取得された文字サイズについての基準出現率Ｓが１０％であったとする。①において、文字サイズが「５」以上のものについての出現率の合計は３％＋５％＝８％であり、この値は上述した基準出現率Ｓを下回っている。一方、文字サイズが「４」以上のものについての出現率の合計は３％＋５％＋７０％＝７８％であり、この値は上述した基準出現率Ｓを超えている。従って、Ｓ１２５の判別処理は、Ｓ１２４の処理において文字サイズが「４」以上のものについての出現率の合計を算出したときにその結果がＹｅｓとなる。そして、このときに実行されるＳ１２６では、その直前のＳ１２４の処理で加算された出現率に対応する文字サイズよりもひとつ大きい文字サイズ、すなわち、「５」が基準文字サイズＥｓｉｚｅとして設定される。従って、後述する索引属性の基準設定処理においては、「５」以上の文字サイズの属性が設定されている文字列、すなわち文字サイズが「５」若しくは「６」に設定されている文字列がＷｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５５】次に、Ｓ１２９の処理によって取得された文字色についての基準出現率Ｃについて、文書１に含まれる文字の文字色についての出現率のうちの最大のものとすることが予め設定されていたとする。②において、出現率が最大の文字色は「黒」であり、その出現率は９０％である。よってＳ１２９の処理ではこの出現率が基準出現率Ｃとして設定されるため、続くＳ１３０及びＳ１３１の処理では、出現率が基準出現率Ｃ以上である文字色、すなわちここでは出現率が９０％以上である黒色

のみが基準色Ｅｃｏｌｏｒとして設定される。従って、後述する索引属性の基準設定処理においては、黒色とは異なる文字色の属性が設定されている文字列、すなわちここでは文字色が「赤」若しくは「青」である文字列がＷｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５６】図８（Ｂ）は、文書２なるＷｅｂページに含まれる文字の出現率を示しており、①はＳ１２２までの処理によって得られた各文字サイズの出現率を示すテーブル、②はＳ１２８までの処理によって得られた各文字色の出現率を示すテーブルである。

【００５７】今、Ｓ１２３の処理によって取得された文字サイズについての基準出現率Ｓが文書１のＷｅｂページと同様に１０％であったとする。①において、文字サイズが「４」以上のものについての出現率の合計は１％であり、この値は上述した基準出現率Ｓを下回っている。一方、文字サイズが「３」以上のものについての出現率の合計は１％＋９９％＝１００％であり、この値は上述した基準出現率Ｓを超えている。従って、Ｓ１２５の判別処理は、Ｓ１２４の処理において文字サイズが「３」以上のものについての出現率の合計を算出したときにその結果がＹｅｓとなる。そして、このときに実行されるＳ１２６では、その直前のＳ１２４の処理で加算された出現率に対応する文字サイズよりもひとつ大きい文字サイズ、すなわち、「４」が基準文字サイズＥｓｉｚｅとして設定される。従って、後述する索引属性の基準設定処理においては、「４」以上の文字サイズの属性が設定されている文字列、すなわち文字サイズが「５」に設定されている文字列がＷｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５８】次に、Ｓ１２９の処理によって取得された文字色についての基準出現率Ｃが１０％であったとする。②において、Ｓ１３０及びＳ１３１の処理では、出現率が基準出現率Ｃ以上である文字色、すなわちここでは出現率が１０％以上である赤色及び青色が基準色Ｅｃｏｌｏｒとして設定される。従って、後述する索引属性の基準設定処理においては、赤色若しくは青色のどちらとも異なる文字色の属性が設定されている文字列、すなわちここでは文字色が「黒」である文字列がＷｅｂページ２０において強調表示を行なっている文字列であるとみなされる。

【００５９】次に、前述した索引生成処理におけるＳ１０９の処理として文字列解析部１２５で行なわれる索引属性の設定処理の詳細について説明する。まず、Ｓ１４１では、ＨＴＭＬフィルタテーブル１２２における、前述したフィルタポインタｎの現在の値によって指定される行に示されている文字列属性が取得される。

【００６０】Ｓ１４２では、前ステップの処理によって取得された文字列属性における文字サイズが、前述した索引属性の基準設定処理によって設定された基準文字サイズＥｓｉｚｅ以上であるか否かが判別され、この判別の結果がＹｅｓのときにのみ、Ｓ１４３において、前述したフィルタポインタｎの現在の値によって指定される行に示されている文字列に対して与えられる属性が一時的に保持される索引属性バッファに定義されている文字サイズフラグが「１」にセットされる。

【００６１】Ｓ１４４では、Ｓ１４１の処理によって取得された文字列属性における文字色が、前述した索引属性の基準設定処理によって設定された基準文字色Ｅｃｏｌｏｒと異なるか否かが判別され、この判別の結果がＹｅｓのときにのみ、すなわち文字色が異なるときにのみ、Ｓ１４５において、前述したフィルタポインタｎの現在の値によって指定される行に示されている文字列に対して与えられる属性が一時的に保持される索引属性バッファに定義されている文字色フラグが「１」にセットされる。

【００６２】Ｓ１４６では、Ｓ１４１の処理によって取得された文字列属性における「太字」の欄にフラグ「１」が格納されているか否かが判別され、この判別の結果がＹｅｓのときにのみ、Ｓ１４７において、前述したフィルタポインタｎの現在の値によって指定される行に示されている文字列に対して与えられる属性が一時的に保持される索引属性バッファに定義されている太字フラグが「１」にセットされる。

【００６３】Ｓ１４６及びＳ１４７の処理を終えるとこの索引属性の設定処理が終了し、前述した図４へと処理が戻る。以上までの処理が索引属性の設定処理である。図３に例示したＷｅｂページ２０に対して以上までに説明した図４の単語索引生成処理が施されることによって生成される索引ファイル３１０のデータ構造を図９に示す。なお、同図においては、図３に示したＷｅｂページ２０の公開されているインターネット４上の論理的な位置を「文書１」として示しているが、例えばこの位置情報をＵＲＬ（Uniform Resource Locator）で示すようにしても勿論構わない。

【００６４】図９における見出し語「簡単」の行を例にして索引ファイル３１０を説明すると、「簡単」の語はインターネット４上における「文書１」なる位置で公開されているＷｅｂページ２０に含まれており、Ｗｅｂページ２０中でこの語は文字サイズによる強調表示、文字色による強調表示、及び太字による強調表示の全てがなされていることが示されている。図３を参照すれば、「簡単」の語について上述した強調表示を行なせることが示されている。

【００６５】次に、情報検索サイト１の有する情報検索管理部２００において行なわれる情報検索の処理の詳細について、具体例を提示して説明する。図１０は情報管理検索部２００で実行される検索処理の処理内容を示すフローチャートである。まず、同図に沿って検索処理の処理内容を説明する。

【００６６】Ｓ２０１では、ブラウザ３０から送られてくる、情報検索の要求及びその検索の対象である単語が示されている検索式が受信されたか否かが情報検索部２１０で判別され、この判別結果がＹｅｓとなったときにのみ、処理がＳ２０２に進む。なお、情報検索部２１０では、検索式が送られてくるとその検索式を検索式格納部２１１に格納する。

【００６７】Ｓ２０２では送られてきた検索式が情報検索部２１０で解析され、その検索式から検索対象である単語が切り出される。Ｓ２０３では、検索単語が切り出された順番で、その検索単語をキーとした索引ファイル３１０の見出し語の検索が情報検索部２１０によって行なわれる。

【００６８】Ｓ２０４では、前ステップでの検索の結果、検索単語に一致する見出し語が発見されたか否かが情報検索部２１０によって判別され、この判別の結果がＹｅｓならばＳ２０５に処理が進み、ＮｏならばＳ２０７に処理が進む。Ｓ２０５では、情報検索部２１０によって、検索単語に一致する見出し語の含まれていた検索レコードから位置情報、要約、及び属性フラグが全て取得され、続くＳ２０６において、この検索単語に一致する見出し語と、位置情報、要約、及び属性フラグとからなるレコードが情報検索部２１０によって検索結果ファイル３２０に格納される。

【００６９】Ｓ２０７では、前述したＳ２０２の処理によって切り出された全ての検索単語について前述したＳ２０３の検索の処理が行なわれたか否かが情報検索部２１０によって判別され、この判別の結果がＹｅｓならばＳ２０８に処理が進み、一方この判別の結果がＮｏならばＳ２０３へと処理が戻って未だ検索の処理の行なわれていない検索単語について上述した処理が行なわれる。

【００７０】ここで、Ｓ２０８において、検索結果ファイル３２０に格納されている位置情報のうち前述したＳ２０２の処理によって切り出された全ての検索単語に該当する位置情報、すなわち検索結果ファイル３２０の全てのレコードに共通に含まれている位置情報が検索結果管理部２２０によって検索結果ファイル３２０から要約と共に抽出される。

【００７１】Ｓ２０９では、前ステップでの処理によって位置情報の抽出が行なえたか否か、すなわち検索結果ファイル３２０の全てのレコードに共通に含まれている位置情報が存在したか否かが検索結果管理部２２０によって判別され、この判別の結果がＹｅｓならばＳ２１０に処理が進み、ＮｏならばＳ２１３に処理が進む。

【００７２】Ｓ２１０では、前述したＳ２０８の処理によって抽出された位置情報と、検索結果ファイル３２０においてその位置情報に対応付けられて格納されている要約とが検索結果管理部２２０によって検索結果リストファイル３３０に格納される。

【００７３】Ｓ２１１では、検索結果ファイル３２０において、前述したＳ２０８の処理によって抽出された位置情報に対応付けられて格納されている属性フラグの個数がその位置情報毎に検索結果管理部２２０によって計数され、この計数結果が属性ポイント数として検索結果リストファイル３３０に格納される。

【００７４】Ｓ２１２では、前ステップによって計数された属性ポイントの計数値の大きい順となるように検索結果リストファイル３３０が検索結果管理部２２０によってソートされる。Ｓ２１３では、検索結果リストファイル３３０に基づき、ソートされた後の検索結果リストの内容をＷｅｂページで表現するＨＴＭＬファイルがＨＴＭＬ作成部４１０によって作成され、続くＳ２１４において、作成されたＨＴＭＬファイルがWWWサーバ部４００によって前述した検索式の送信元であるブラウザ３０へ宛てて送出され、この検索処理が終了する。

【００７５】以上までの処理が検索処理である。次に、この検索処理について、索引ファイル３１０として図９に示したものが情報データベース管理部３００に格納されている場合を例にして説明する。まず、ブラウザ３０から検索式が送られてくると、Ｓ２０１の判別の結果がＹｅｓとなり、続くＳ２０２において検索単語の切り出しが行なわれる。ここでは、このＳ２０２の処理の結果、検索単語として「ハンバーグ」、「トマト」の各語が切り出されたものとする。

【００７６】検索単語の切り出しが完了すると処理はＳ２０３に進み、まず、索引ファイル３１０の見出し語について検索単語「ハンバーグ」の検索が行なわれ、図９における見出し語「ハンバーグ」についてのレコードが発見される。従ってＳ２０４の判別処理の結果はＹｅｓとなり、Ｓ２０５に処理が進む。

【００７７】Ｓ２０５では発見されたレコードから位置情報、要約、及び属性フラグが全て取得され、続くＳ２０６においてその位置情報、要約、及び属性フラグからなるレコードが検索結果ファイル３２０に格納される。その後、Ｓ２０７における判別処理が行なわれるが、前述したＳ２０２の処理によって切り出された検索単語のうち「トマト」についてはＳ２０３の検索の処理が未だ行なわれていないので、Ｓ２０７の判別処理の結果はＮｏとなり、処理はＳ２０３へと戻る。

【００７８】以降、検索単語「トマト」について上述した検索単語「ハンバーグ」と同様の処理が行なわれ、図９に示す索引ファイル３１０から見出し語「トマト」についてのレコードが発見されてそのレコードにおける位置情報、文字列、及びリンクフラグと検索単語「トマト」とからなるレコードが検索結果ファイル３２０に格納される。

【００７９】以上までの処理によって生成される検索結果ファイル３２０の内容を図１１に示す。この図１１に示す検索結果ファイル３２０が生成された後にはＳ２０７の判別処理の結果がＮｏとなり、処理はＳ２０８に進

む。

【００８０】Ｓ２０８では、検索結果ファイル３２０の全てのレコードに共通に含まれている位置情報の抽出が行なわれ、その結果、「ハンバーグ」及び「トマト」の両レコードに共通に含まれている位置情報としてここでは「文書１」、「文書２」、及び「文書３」の３つが抽出されとものとする。従って、続くＳ２０９の判別処理の結果はＹｅｓとなり、処理はＳ２１０に進む。

【００８１】Ｓ２１０では、抽出された「文書１」、「文書２」、及び「文書３」の３つの位置情報と、検索結果ファイル３２０においてその位置情報に対応付けられて格納されている文字列とが検索結果リストファイル３３０に格納され、続くＳ２１１において、抽出された位置情報「文書１」、「文書２」、及び「文書３」の各々に対応付けられて格納されている属性フラグの個数がそれぞれ計数され、その計数結果が属性ポイント数として検索結果リストファイル３３０に格納される。

【００８２】ここで図１２について説明する。同図は、検索結果リストファイル３３０の内容を示しており、上述したＳ２１１までの処理によって、同図（ａ）に示すものが検索結果リストファイル３３０として作成される。図１１に示す検索結果ファイル３２０では、「文書１」についての属性フラグは合計６つ格納されているので、図１２（ａ）に示す検索結果リストファイル３３０における位置情報「文書１」についての属性ポイント数は「６」とされる。

【００８３】「文書２」及び「文書３」の属性ポイントについても同様であり、図１１に示す検索結果ファイル３２０より、図１２（ａ）に示す検索結果リストファイル３３０における「文書２」についての属性ポイント数は「０」とされ、「文書３」についての属性ポイント数は「３」とされる。

【００８４】前述したＳ２１１までの処理によって図１２（ａ）にその内容を示す検索結果リストファイル３３０が作成されると、続くＳ２１２において、属性ポイント数値の大きい順となるように検索結果リストファイル３３０のソートが行なわれる。図１２（ａ）の検索結果リストファイル３３０に対して属性ポイント数に基づくソートの行なわれた結果が図１２（ｂ）に示されているものであり、各行の順番が属性ポイントの高い「文書１」、「文書３」、「文書２」の順に並べ替えられている。

【００８５】その後、Ｓ２１３において、図１２（ｂ）のようにソートが行なわれた検索結果リストファイル３３０の内容をＷｅｂページで表現するＨＴＭＬファイルが作成され、続くＳ２１４において作成されたＨＴＭＬファイルが送出されてこの検索処理が終了する。

【００８６】作成されたＨＴＭＬファイルがブラウザ３０によって閲覧されることによって表示される、情報検索の結果を示すＷｅｂページの画面例を図１３に示す。同図に示す画面において、「ハンバーグ」及び「トマト」の語についての検索結果であって重要度の高い情報の得られることの期待されるものから優先的に並べられている、「文書１」、「文書３」及び「文書２」の位置情報には各々その位置情報で示されるＷｅｂページ２０へのハイパーリンクが埋め込まれ、この検索結果の利用者への便宜が図られている。

【００８７】なお、以上までに説明した実施形態におけるＷｅｂページ２０についてのＨＴＭＬソースの解析では、その文字列がＷｅｂページ２０中で強調されているかどうかを、<;FONT>;タグ及び<;B>; タグの記述に基づいて判定しているが、この他のタグの記述に基づいてこの判定を行なうようにしてもよい。この強調の判定に採用することのできるタグの例としては、文字列を斜体文字で表示させる<;I>; タグや文字列に下線を付す<;U>; タグ、あるいは標準的なブラウザでは単に太字文字で文字列を表示させるに過ぎないもののＷｅｂページに記載されている文章を音声により読み上げる音声ブラウザではその文字列を強く発音させるようにすることのできる<;STRONG>;タグなどがある。また、文字列の表示に使用されるフォントの種類を指定するためのFACE属性が指定されている<;FONT>;タグに基づき、そのＷｅｂページの表示ために通常使用されるものとは異なるフォントが指定されている文字列はそのＷｅｂページにおいて強調されているものと判定するようにしてもよい。

【００８８】なお、以上までに説明した本発明の実施形態において情報サイト１が行なっていた索引生成処理及び検索処理と同様の処理を前述したような標準的な構成を有するコンピュータに行なわせるための制御プログラムを作成し、その制御プログラムをそのコンピュータに読み込ませて実行させることにより、このようなコンピュータで本発明を実施することができる。

【００８９】また、このような制御プログラムをコンピュータで読み取り可能な記録媒体に記録させ、そのプログラムを記録媒体からコンピュータに読み出させて実行させることによって本発明をコンピュータで実施することも可能である。記録させた制御プログラムをコンピュータで読み取ることの可能な記録媒体の例を図１４に示す。同図に示すように、記録媒体としては、例えば、コンピュータ５０１に内蔵若しくは外付けの付属装置として備えられるＲＯＭやハードディスク装置などの記憶装置５０２、あるいはフレキシブルディスク、ＭＯ（光磁気ディスク）、ＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭなどといった携帯可能記録媒体５０３等が利用できる。また、記録媒体はネットワーク５０４を介してコンピュータ５０１と接続される、プログラムサーバ５０５として機能するコンピュータが備えている記憶装置５０６であってもよい。この場合には、制御プログラムを表現するデータ信号で搬送波を変調して得られる伝送信号を、プログラムサーバ５０５５から伝送媒体であるネットワーク５０

４を通じて伝送するようにし、コンピュータ５０１では受信した伝送信号を復調して制御プログラムを再生することで当該制御プログラムを実行できるようになる。

【００９０】

【発明の効果】本発明によれば、通信ネットワーク上で公開されている文書情報に含まれている文字列を構成する単語に、その文書情報の位置を示す位置情報、及びその文字列に与えられている強調を示す強調属性を対応付けて索引ファイルに登録することで、検索対象を表す単語に基づいてその索引ファイルの検索を行った場合に、その検索によって取得された位置情報のうち、その位置情報に対応付けられた単語に強調属性が対応付けられているものを優先して提示することができるので、検索目的に対してより適切な情報検索結果を情報検索者に提供することができる。

【図面の簡単な説明】

【図１】本発明を実施する情報検索サイトが情報検索サービスを提供する通信ネットテワークの全体構成を示す図である。

【図２】情報検索サイトの詳細構成を示す図である。

【図３】Ｗｅｂページの一例を示す図である。

【図４】単語索引生成処理の処理内容を示すフローチャートである。

【図５】ＨＴＭＬフィルタテーブルの例を示す図である。

【図６】索引属性の基準設定処理の処理内容を示すフローチャートである。

【図７】索引属性の設定処理の処理内容を示すフローチャートである。

【図８】索引属性の設定処理を説明する図である。

【図９】索引ファイルのデータ構造を示す図である。

【図１０】検索処理の処理内容を示すフローチャートである。

【図１１】索引結果ファイルの例を示す図である。

【図１２】索引結果リストファイルのソートの様子を示す図である。

【図１３】情報検索の結果を示すＷｅｂページの画面例を示す図である。

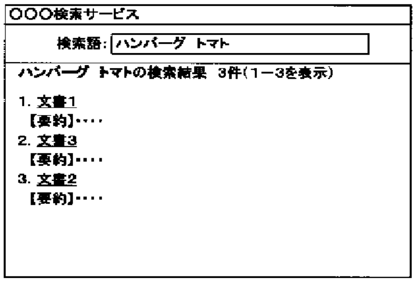
【図１４】記録させたプログラムをコンピュータで読み取ることの可能な記録媒体の例を示す図である。

【符号の説明】

１	情報検索サイト
２ａ、２ｂ、２ｃ、２ｄ	情報提供サイト
３ａ、３ｂ	ユーザ端末
４	インターネット
２０、２０ａ、２０ｂ、２０ｃ、２０ｄ	Ｗｅｂページ
３０、３０ａ、３０ｂ	ブラウザ
１００	情報管理部
１１０	Ｗｅｂページ収集管理部
１２０	索引作成管理部
１２１	Ｗｅｂページ解析部
１２２	ＨＴＭＬフィルタテーブル
１２３	単語抽出管理部
１２４	文字強調解析部
１２５	文字列解析部
１２６	索引登録部
２００	情報検索管理部
２１０	情報検索部
２１１	検索式格納部
２２０	検索結果管理部
３００	データベース管理部
３１０	索引ファイル
３２０	検索結果ファイル
３３０	検索結果リストファイル
４００	WWWサーバ部
４１０	ＨＴＭＬ作成部
５０１	コンピュータ
５０２、５０６	記憶装置
５０３	携帯可能記録媒体
５０４	ネットワーク
５０５	プログラムサーバ

【図１３】

情報検索の結果を示すWebページの画面例を示す図



〇〇〇検索サービス

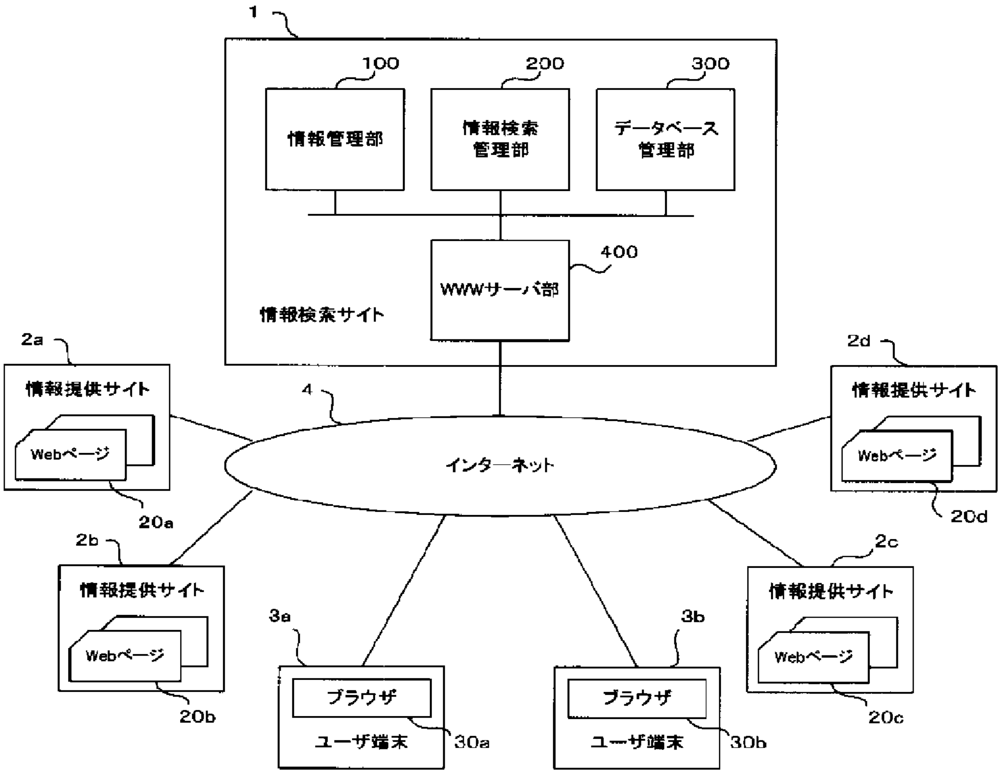
検索語: ハンバーグ トマト

ハンバーグ トマトの検索結果　3件(1－3を表示)

1. 文書1
【要約】････
2. 文書3
【要約】････
3. 文書2
【要約】････

【図１】

本発明を実施する情報検索サイトが情報検索サービスを提供する通信ネットワークの全体構成を示す図



【図１１】

索引結果ファイルの例を示す図

見出し語	位置情報1	要約	属性フラグ1			位置情報2	要約	属性フラグ2			位置情報3	要約	属性フラグ3		
			文字サイズ	色	太字			文字サイズ	色	太字			文字サイズ	色	太字
ハンバーグ	文書1	…	1	1	1	文書2	…	0	0	0	文書3	…	0	1	0
トマト	文書1	…	1	1	1	文書2	…	0	0	0	文書3	…	1	0	1

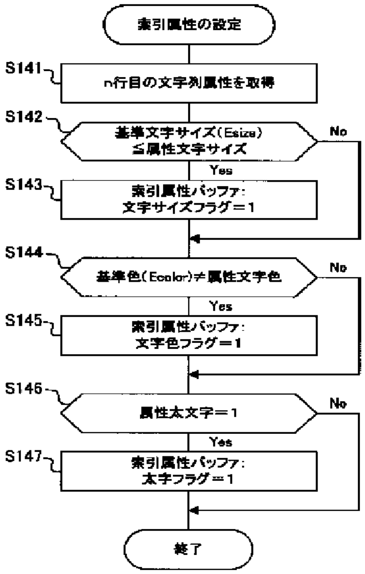
【図５】

HTMLフィルタテーブルの例を示す図

種別	文字列属性			文字列
	文字サイズ	色	太字	
STRING	6	赤	1	簡単料理
BR				
STRING	5	青	1	◇ハンバーグのトマトソース煮
BR				
STRING	4	黒		【材料】4人分
BR				
STRING	4	黒		冷凍ハンバーグ：4個
BR				
・・・	・・・	・・・	・・・	・・・
STRING	4	黒		【作り方】
BR				
STRING	4	黒		1．・・・
BR				
・・・	・・・	・・・	・・・	・・・

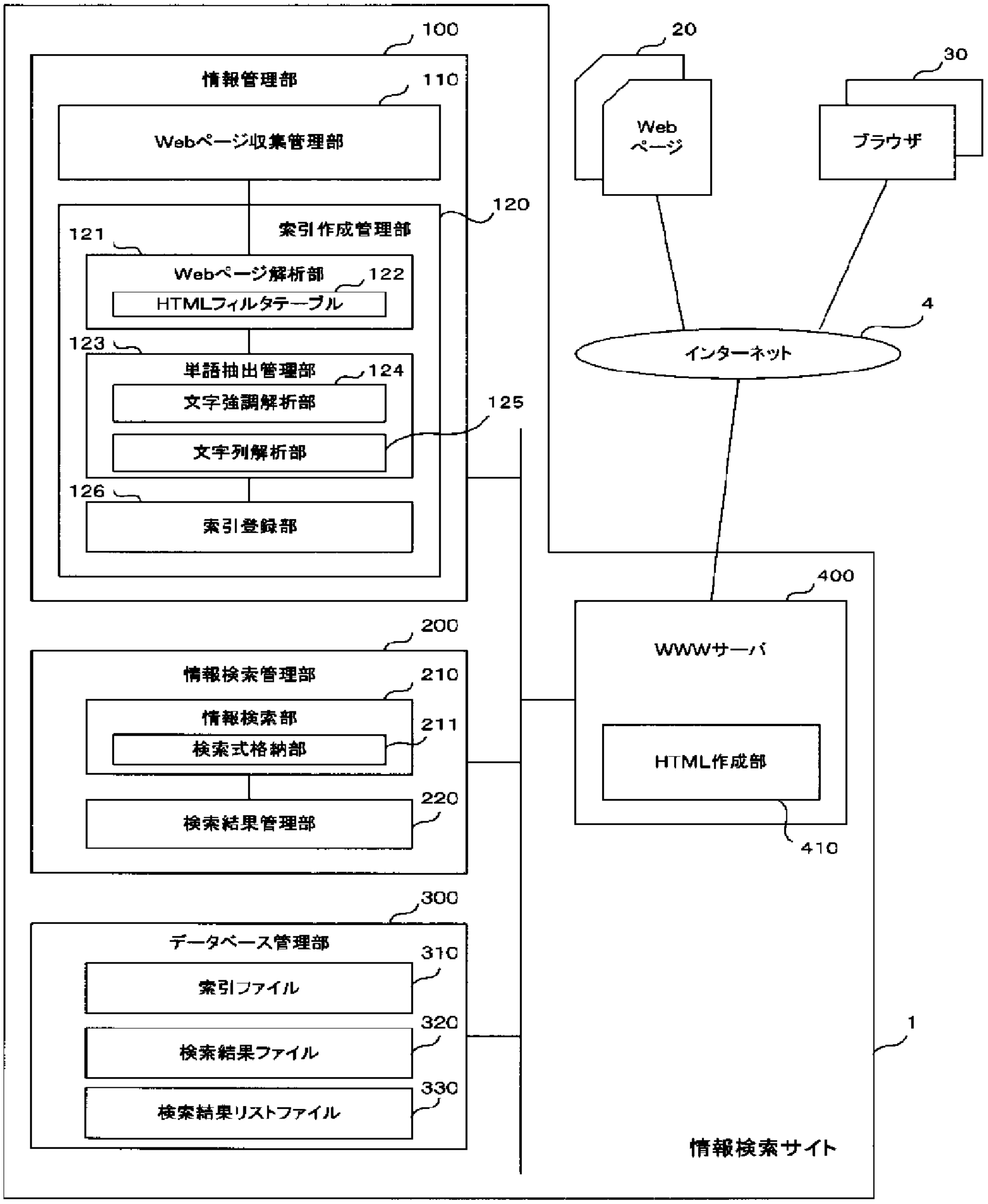
【図７】

索引属性の設定処理の処理内容を示すフローチャート


索引属性の設定
S141
n行目の文字列属性を取得
S142
基準文字サイズ（Esize）≦属性文字サイズ
No
Yes
S143
索引属性バッファ：文字サイズフラグ＝1
S144
基準色（Ecolor）≠属性文字色
No
Yes
S145
索引属性バッファ：文字色フラグ＝1
S146
属性太文字＝1
No
Yes
S147
索引属性バッファ：太字フラグ＝1
終了

【図２】

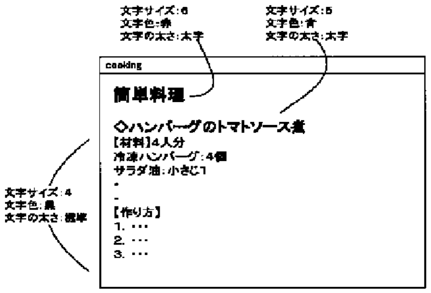
情報検索サイトの詳細構成を示す図



【図３】

Webページの一例を示す図

（a）ブラウザによって表示されるWebページ画面



cooking

簡単料理

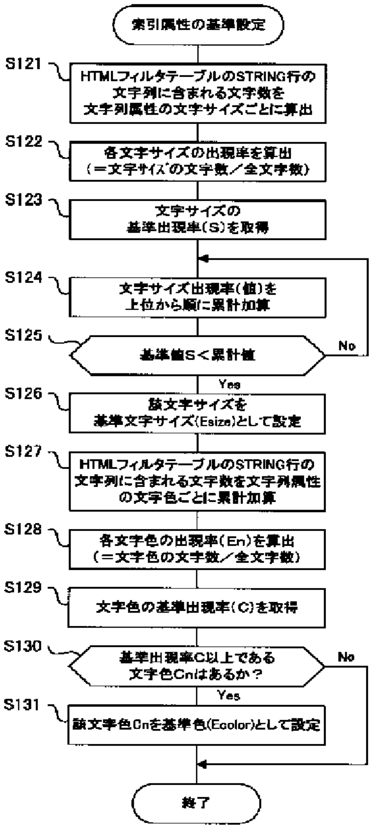
◇ハンバーグのトマトソース煮
【材料】4人分
冷凍ハンバーグ:4個
サラダ油:小さじ1
・
・
【作り方】
1. ･･･
2. ･･･
3. ･･･

（b）HTMLソース

```
<HTML>
<HEAD>
 <TITLE>cooking</TITLE>
</HEAD>
<BODY>
<FONT SIZE="6"COLOR="#FF0000"><B>簡単料理</B></FONT><BR>
<P> </P>
<FONT SIZE="5"COLOR="#0000FF"><B>◇ハンバーグのトマトソース煮
</B></FONT><BR>
【材料】4人分<BR>
冷凍ハンバーグ:4個<BR>
サラダ油:小さじ1<BR>
   ・
   ・
【作り方】<BR>
1. ･･･<BR>
2. ･･･<BR>
3. ･･･<BR>
   ・
   ・
</BODY>
</HTML>
```

【図６】

索引属性の基準設定処理の処理内容を示すフローチャート


索引属性の基準設定
S121
HTMLフィルタテーブルのSTRING行の文字列に含まれる文字数を文字列属性の文字サイズごとに算出
S122
各文字サイズの出現率を算出（＝文字サイズの文字数／全文字数）
S123
文字サイズの基準出現率(S)を取得
S124
文字サイズ出現率(値)を上位から順に累計加算
S125
基準値S＜累計値
No
Yes
S126
該文字サイズを基準文字サイズ(Esize)として設定
S127
HTMLフィルタテーブルのSTRING行の文字列に含まれる文字数を文字列属性の文字色ごとに累計加算
S128
各文字色の出現率(En)を算出（＝文字色の文字数／全文字数）
S129
文字色の基準出現率(C)を取得
S130
基準出現率C以上である文字色Cnはあるか？
No
Yes
S131
該文字色Cnを基準色(Ecolor)として設定
終了

【図１２】

検索結果リストファイルのソートの様子を示す図

（a）ソート前

位置情報1	要約	ポイント数
文書1	…	6
文書2	…	0
文書3	…	3

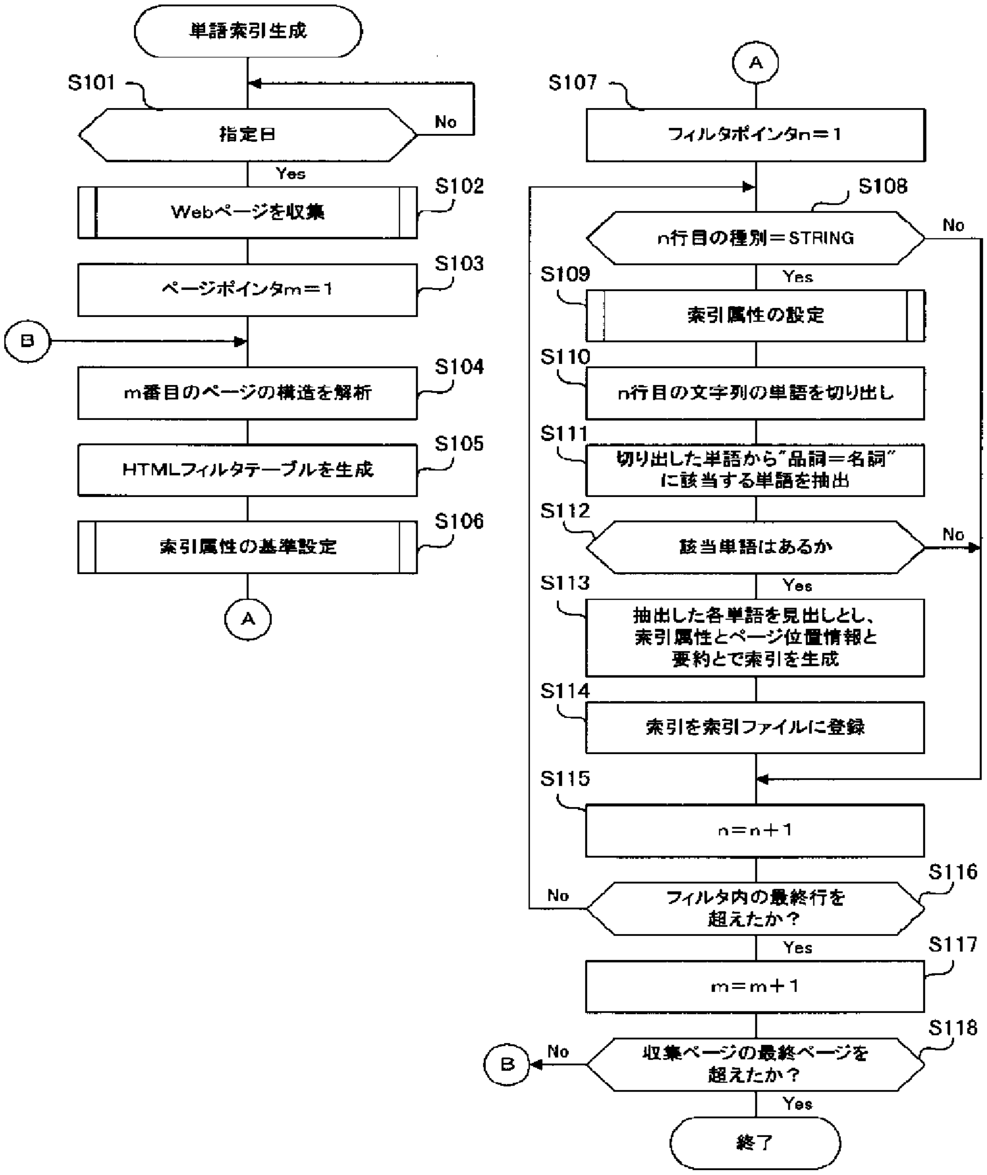
↑ソート

（b）ソート後

位置情報1	要約	ポイント数
文書1	…	6
文書3	…	3
文書2	…	0

【図４】

単語索引生成処理の処理内容を示すフローチャート



【図８】

索引属性の設定処理を説明する図

(A)文書1

①ページ内文字サイズ出現率テーブル

基準文字サイズ
＝最大10%以上

文字サイズ	出現率
7	—
6	3%
5	5%
4	70%
3	12%
2	10%
1	—

②ページ内文字色出現率テーブル

最大出現率文字色
（＝90%以上）以外
の文字色

基準色

文字色	出現率
赤	4%
青	6%
黒	90%

(B)文書2

①ページ内文字サイズ出現率テーブル

基準文字サイズ
＝最大10%以上

文字サイズ	出現率
7	—
6	—
5	1%
4	—
3	99%
2	—
1	—

②ページ内文字色出現率テーブル

出現率
10%以上

文字色	出現率
赤	50%
青	45%
黒	5%

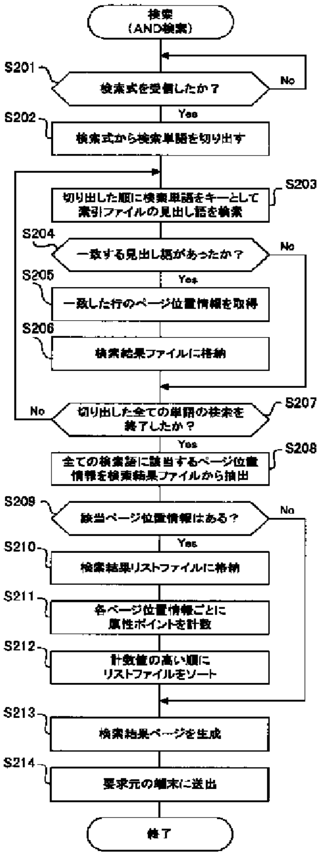
【図９】

索引ファイルのデータ構造を示す図

見出し語	位置情報1	要約	属性フラグ1			位置情報2	要約	属性フラグ2			…
			文字サイズ	色	太字			文字サイズ	色	太字	
…	…	…	..	..	..	…	…	..	..	..	…
簡単	文書1	…	1	1	1						…
料理	文書1	…	1	1	1	文書2	…	0	0	0	…
ハンバーグ	文書1	…	1	1	1	文書2	…	0	0	0	…
トマト	文書1	…	1	1	1	文書2	…	0	0	0	…
…	…	…	..	..	..	…	…	..	..	..	…
材料	文書1	…	0	0	0						…
人分	文書1	…	0	0	0						…
…	…	…	..	..	..	…	…	..	..	..	…
日記	文書2	…	1	1	1						…
…	…	…	..	..	..	…	…	..	..	..	…

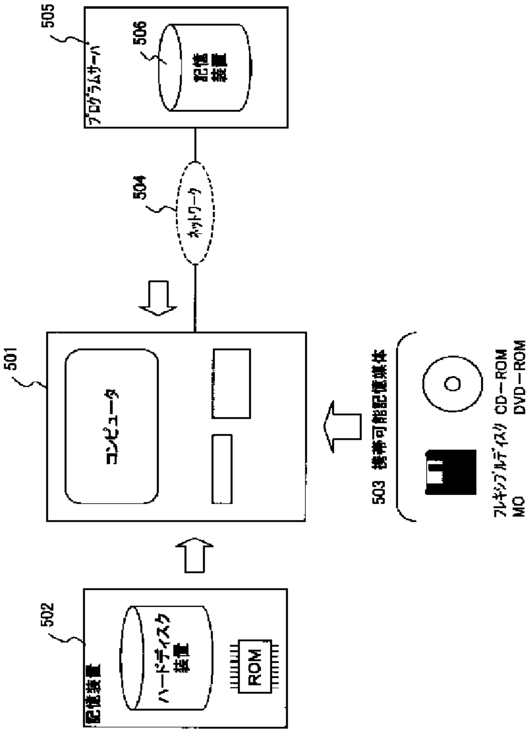
【図１０】

図10の検索処理の処理内容を示すフローチャート


検索
（AND検索）
S201
検索式を受信したか？
No
Yes
S202
検索式から検索単語を切り出す
S203
切り出した順に検索単語をキーとして
索引ファイルの見出し語を検索
S204
一致する見出し語があったか？
No
Yes
S205
一致した行のページ位置情報を取得
S206
検索結果ファイルに格納
S207
No
切り出した全ての単語の検索を
終了したか？
Yes
S208
全ての検索語に該当するページ位置
情報を検索結果ファイルから抽出
S209
該当ページ位置情報はある？
No
Yes
S210
検索結果リストファイルに格納
S211
各ページ位置情報ごとに
属性ポイントを計数
S212
計数値の高い順に
リストファイルをソート
S213
検索結果ページを生成
S214
要求元の端末に送出
終了

【図１４】

記録させた制御プログラムをコンピュータで
読み取ることの可能な記録媒体の例を示す図


505
プログラムサーバ
506
記憶装置
504
ネットワーク
501
コンピュータ
503 携帯可能記憶媒体
フレキシブルディスク
MO
CD−ROM
DVD−ROM
502
記憶装置
ハードディスク装置
ROM